Journal of Building Performance Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tbps20

A hierarchical Bayesian framework for calibrating micro-level models with macro-level data

A.T. Booth [a], R. Choudhary [a] & D.J. Spiegelhalter [b]

[a] Energy Efficient Cities Initiative, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, UK

[b] Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK

Published online: 31 Jan 2013.

Journal of Building Performance Simulation, 2013
Vol. 6, No. 4, 293–318, http://dx.doi.org/10.1080/19401493.2012.723750

A hierarchical Bayesian framework for calibrating micro-level models with macro-level data

A.T. Booth[a]*, R. Choudhary[a] and D.J. Spiegelhalter[b]

[a]*Energy Efficient Cities Initiative, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK;* [b]*Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK*

(*Received 20 February 2012; final version received 20 August 2012*)

Owners of housing stocks require reliable and flexible tools to assess the impact of retrofits technologies. Bottom-up engineering-based housing stock models can help to serve such a function. These models require calibrating, using micro-level energy measurements at the building level, to improve model accuracy; however, the only publicly available data for the UK housing stock is at the macro-level, at the district, urban, or national scale. This paper outlines a method for using macro-level data to calibrate micro-level models. A hierarchical framework is proposed, utilizing a combination of regression analysis and Bayesian inference. The result is a Bayesian regression method that generates estimates of the average energy use for different dwelling types whilst quantifying uncertainty in both the empirical data and the generated energy estimates. Finally, the Bayesian regression method is validated and the use of the hierarchical Bayesian calibration framework is demonstrated.

Keywords: Bayesian; calibration; regression; retrofit; housing stock; uncertainty

1. Introduction

1.1. Calibration

Housing stock energy models come in a variety of forms, as described in Swan and Ugursal (2009). Such models can be broadly divided into two camps: 'top-down' and 'bottom-up'. Top-down models tend to be of a statistical nature, attributing the energy consumption of the residential sector at a macro-level – i.e. at an urban or national scale – to macroeconomic factors (such as the price of energy), climatic conditions (such as the external temperature), or technological characteristics of the housing stock (such as the average boiler efficiency). Bottom-up models, in contrast, attempt to calculate energy consumption at a micro-level – i.e. at the level of an individual dwelling – and then aggregate or extrapolate in order to estimate the energy consumption at a macro-level. Bottom-up models can be further sub-divided into those that use statistical methods, such regression or neural network analysis, and those that are engineering-based, using physical equations to calculate end-use energy demands.

The choice of model type will depend on the scale of the scenario that is to be analyzed and the data sources that are available. Top-down models are more suitable in cases where the user is interested in the response of a housing stock to long-term, large-scale changes. In addition, top-down models utilize aggregated, macro-level data, such as national energy statistics or average physical characteristics of the stock. This macro-level data is often more easily available than the micro-level data, such as utility bills and technical surveys of individual dwellings, which must be gathered for a representative sample of the housing stock in order to construct an accurate bottom-up model.[1]

Engineering-based housing stock energy models, such as those reviewed by Kavgic *et al.* (2010), are most suitable for providing decision makers with a cost-benefit analysis of different retrofit options. Unlike top-down or bottom-up statistical models, engineering-based methods attempt to model the actual physical characteristics of a building and its systems, and thus allow the effects of different technologies to be examined. However, in order to improve the accuracy of engineering-based housing stock models, calibration of the input parameters is required. Calibration should focus on the most sensitive input parameters – i.e. those parameters for which a change in the input produces the largest change in the output. For housing stock energy models, these have been found by Firth *et al.* (2010), Cheng and Steemers (2011), and Booth *et al.* (2012) to include parameters that describe the physical characteristics of the building and its systems, such as the U-value of glazing, the air leakage (infiltration) rate, and the efficiency of the heating system, as well as parameters that describe how

*Corresponding author. Email: atb29@cam.ac.uk

the dwelling is used, such as the internal set point temperature and number of hours for heating.

In addition, the uncertainty associated with these inputs – along with other sources of uncertainty, as described in Spiegelhalter and Best (2003) and Spiegelhalter and Riesch (2011) – should be accounted for and quantified in order to display risks to decision makers. Bayesian calibration methods, e.g. Kennedy and O'Hagan (2001), can be used to incorporate such uncertainties into the calibration process.

Calibration is achieved by comparing model outputs against measured data. This is often done via an iterative procedure of trial-and-error. The use of Bayesian methods for calibrating engineering-based building energy models has been demonstrated by the authors of this paper (see Booth *et al.* 2012) and by other authors (e.g. Heo *et al.* 2012) in previous literature. A Bayesian approach enables quantification of uncertainties in input parameters, which can be propagated through the housing stock model using probabilistic sensitivity analysis, as described in Lomas and Eppel (1992) and Macdonald (2002). The probabilistic outcomes of the model naturally allow quantification of relative risk of different retrofit technologies in meeting energy saving targets.

In addition, Bayesian methods reduce the amount of empirical data that is needed for calibration by utilizing prior knowledge, such as expert opinion; nevertheless, some empirical data is still needed. Specifically, data is needed that is of the same resolution as the output of the model. For example, calibration of bottom-up engineering-based models that calculate energy demand at the building level will require energy consumption data for individual buildings.

1.2. Extrapolation to macro-level analysis

In a recent study (Booth *et al.* 2012), the use of Bayesian methods was demonstrated for calibrating an engineering-based energy model of a group of flats, using empirical energy data that was collected on a daily basis for each dwelling. In addition, Booth *et al.* (2012) proposed to develop the Stochastic Urban Scale Domestic Energy Model (SUSDEM) for probabilistic retrofit analysis of an entire housing stock comprising of diverse sets of dwelling types. However, in order to extend the proposed methodology underlying SUSDEM to analyze an entire housing stock, empirical energy data is needed that covers a representative sample of all types of dwellings within the housing stock.

Such high-resolution, micro-level empirical data, however, is difficult to obtain for a representative sample of the UK housing stock: energy data are not publically available for individual dwellings; utility companies are restricted in what information that can provide at the micro-level due to confidentially; conducting a longitudinal study of energy use for a sample of dwellings that is representative of the housing stock is expensive and time consuming, and cannot be easily updated to represent shifts in consumption. Indeed, a review of bottom-up engineering-based residential energy models in Kavgic *et al.* (2010) concludes that the lack of publicly available energy consumption data is one of the main barriers to the future development of robust housing stock models.

Whilst energy consumption data is not easily obtainable at the micro-level, i.e. at the level of individual buildings, publicly available energy data is available in the UK at the macro-level, i.e. at an aggregated level for hundreds of dwellings. In this paper, we examine ways of leveraging the availability of annually updated aggregated energy data to calibrate engineering-based energy models of representative dwelling types within the housing stock. We propose a hierarchical framework in which a top-down statistical model is used to infer the energy consumption of each representative dwelling type from aggregated energy consumption statistics available publically. The distribution of energy consumption per building type inferred through this top-down analysis serves the role of 'data' for calibrating the input parameters of the bottom-up probabilistic engineering models. A Bayesian regression method is employed for the top-down statistical model in order to account for uncertainties in the macro-level data and the calculated estimates of energy consumption. Meanwhile, Bayesian calibration is suggested to improve the accuracy of the micro-level, engineering-based model.

The strength of Bayesian methods, in this context, lies in their ability to take uncertainties into account and to combine multiple sources of information of varying scales and reliability. This is ideal in the case of a housing stock analysis, where data is continually being gathered by a variety of agents at a variety of scales. The use of a Bayesian hierarchical framework, therefore, allows analysis to be conducted flexibly depending on what information is available to users: if energy data is only available at the macro-level, then this can be used within the hierarchical framework to constrain the outcomes from a bottom-up engineering-based model for retrofit analysis; as energy consumption data becomes available at the micro-level – i.e. from monitoring individual dwellings – the uncertainties in the model outcomes can be further reduced with this additional information.

The focus of this paper is to outline the proposed top-down stochastic method for inferring energy consumption per dwelling type using aggregated macro-level energy statistics. The main sources of macro-level data are outlined in section 2.1, whilst the top-down stochastic method is described in section 2.2. An illustrative study is described in section 2.3, the results of which are reported in section 3.1. The limitations of using such an approach also addressed: the integration of multiple sources of data into a single analysis is examined in section 3.2; quantification of uncertainty in the data itself is explored in section 3.3. A validation of the results obtained from the case study is provided in section 4.1, whilst the synthesis of a top-down model and bottom-up engineering models within a hierarchical framework is discussed and demonstrated in section 4.2.

2. Method

2.1. Data sources

Macro-level data is publicly available for residential energy consumption in the UK in the form of domestic gas and electricity consumption data from the Department of Energy and Climate Change (DECC).[2] This data is available at two different resolutions: for the middle layer super output area (MLSOA) and the lower level super output area (LLSOA). The MLSOA and the LLSOA are classifications used by the Office for National Statistics (ONS) to define an area, based upon an equal population: each MLSOA has a minimum population of approximately 5000 with an average of approximately 7200, whilst each LLSOA has a minimum population of approximately 1000, with an average of approximately 1500. This equates to an average of approximately 3000 households for each MLSOA and 600 households for each LLSOA. In this study, we focus on the use of LLSOA data, since it is the highest resolution data that is publicly available.

Dimensions of dwellings can be obtained from geographic information and mapping sources, which include the footprint area of a building as well as its height, which is obtained using LIDAR[3] technology. In addition, building class data for each dwelling can also be obtained, which includes information on the building type (i.e. semi-detached, detached, terraced, flat, etc.) and the building age (i.e. age band of construction) from architectural surveys. In this project, these two data sources of the building dimensions and the building class are combined into one database using geospatial information system (GIS) software.

2.2. Top-level model formulation

For the top-level model formulation, a linear relationship can be used to represent the total energy consumption for a set of LLSOAs ($\mathbf{E} = [E_1, E_2, \ldots, E_N]$) as a function of the average energy consumption of a building class per m^2 ($\mathbf{e} = [e_1, e_2, \ldots, e_M]$), and the total floor areas for the building classes in the LLSOAs ($\mathbf{X}$):

$$\mathbf{E} = \mathbf{eX} + \boldsymbol{\delta} \tag{1}$$

where

$$E_i = \sum_{j=1}^{P}(e_j x_{ij}) + \delta_i \tag{2}$$

In Equation (2), E_i is the known energy consumption (in kWh/year) for the ith LLSOA, x_{ij} is the known total floor area (in m^2) for building class j in LLSOA i, and e_j is the unknown average energy intensity (in kWh/m^2/year) for a building class j. $\boldsymbol{\delta}$ is a random error term that takes into account the difference between the macro-level energy data, $\mathbf{E}$, and the predictions from the linear approximation, $\mathbf{eX}$ arising due to measurement errors in the data.

The energy data (E_i) and floor areas (x_{ij}) were both normalized by the total floor area for each LLSOA ($\sum_{j=1}^{P} x_{ij}$) for this study in order to improve comparison between the empirical energy data and the energy use intensity per building class. The normalized energy data was therefore in units of kWh/m^2/year, and the normalized floor areas are dimensionless variables representing the fraction of the total floor area for each LLSOA that a specific building class represents.

It should be noted that the linear equations shown in Equations (1) and (2) are not linear approximations of the complex, non-linear processes that determine a building's energy consumption. Instead, Equations (1) and (2) simply show a mathematical relationship that expresses the energy consumption at a district level in terms of the average energy intensities of different building classes (in terms of kWh/m^2/year) and the total floor area of these building classes. If the number of available data points for energy consumption (N) is equal to or greater than the number of defined building classes (P), i.e. $N > P$, then a linear regression model (such as ordinary least squares) can be used to infer the average energy intensity for a building class ($\mathbf{e} = [e_1, e_2, \ldots, e_M]$) from known values of $\mathbf{E}$ and $\mathbf{X}$.

A linear regression model, however, has two main flaws. Firstly, the energy intensity estimates (i.e. the unknown regression coefficients) are not constrained, and can therefore result impossible values (i.e. $e_j < 0$) or highly improbable values (i.e. very small or extremely large energy intensity estimates). Secondly, linear regression models do not accurately quantify the uncertainty that is involved in the analysis. For example, the OLS method returns a singular set of estimated values for the regression coefficients ($\mathbf{e}$) along with confidence intervals, rather than providing a distribution of possible values that quantifies the probability of a set of energy intensity estimates given the empirical energy consumption data, **E**, and measured building class areas, **X**. In addition, there is uncertainty over the accuracy of the empirical energy consumption data itself, due to the method by which it is collected, which should also be quantified accurately.

These flaws can be addressed using a Bayesian approach, which incorporates prior knowledge and the effects of uncertainty into the regression analysis, as demonstrated in Caves *et al.* (1987). With a Bayesian regression analysis, impossible or highly improbable estimates of energy intensities can be avoided by specifying prior distributions for the energy intensity of each building class e_j that restricts the possible outcome of the regression analysis.

Not only does the use of prior distributions eliminate the possibility of obtaining impossible values for energy intensity estimates, but they also allow for prior beliefs and expert judgement to be incorporated into the regression analysis along with the empirical data. These prior beliefs and expert opinions may include sources such as: incomplete sets of previous data or small sub-samples; empirical data relating to similar, but not identical, scenarios or experiments; literature review; benchmarks and standards; 'rules-of-thumb'. A process known as 'elicitation' (O'Hagan 2006) is used to turn this partial information from prior sources and expert opinion into a quantifiable probability distribution (i.e. a prior distribution). If the prior knowledge is very strong (i.e. a high level of confidence in the expert opinion), then a narrow prior distribution is used and the results of the regression analysis will tend towards the prior values provided. Likewise, if the prior knowledge is very weak or non-existent, then a wide prior distribution (e.g. a wide uniform distribution) is used and the regression analysis will be dominated by the empirical data; the results will then tend towards those of the standard linear regression.

The Bayesian approach also allows for uncertainty over the empirical energy data to be taken into account and accurately quantified. The notation for the Bayesian regression model is shown below:

$$E_i \sim N(\mu_i, \sigma) \tag{3a}$$

$$\mu_i = \sum_{j=1}^{P} (e_j x_{ij}) \tag{3b}$$

The empirical energy data is assumed to be stochastically drawn from a normal distribution, as shown in Equation (3a), with a mean value that is obtained from the regression model (μ_i) given in Equation (3b) and an unknown standard deviation σ. The use of a normal distribution for E_i in Equation (3a) assumes that any measurement error associated with the empirical energy data is unbiased.

Bayesian inferences involve the use of Bayes' theorem, as shown in Equation (4), whereby empirical data (Z) is used to adjust the probability distributions (i.e. priors) assigned to uncertain parameters (θ):

$$\Pr(\theta \mid Z) = \frac{\Pr(Z \mid \theta)\Pr(\theta)}{\Pr(Z)} \tag{4}$$

In Equation (4), $\Pr(\theta)$ is the (joint) prior distribution of the uncertain parameters, which is updated based on the calculated likelihood of the empirical data given the uncertain parameters – $\Pr(Z \mid \theta)$ – in order to calculate the (joint) posterior distribution, $\Pr(Z \mid \theta)$. In the case of this study, the empirical data is the energy consumption data (**E**) and the uncertain parameters of interest are the energy intensities (**e**). The likelihood function, therefore, is given by the linear regression formulation shown in Equation (3a).

For very simple examples, Bayesian inference can be conducted by hand, generating an exact parametric expression for the posterior distribution. For most problems, however, the posterior distribution must be calculated through a process of sampling, using Monte Carlo techniques. For this study, the Bayesian inference was conducted using the Bayesian modelling software WinBUGS[4] (see Lunn *et al.* 2000), which is based upon Markov Chain Monte Carlo (MCMC) methods. WinBUGS works by allowing the user to specify statistical relationships between different variables, either probabilistically or deterministically, in order to generate a joint conditional (posterior) distribution over all variables. Empirical data can be specified for any observed variable and prior distributions are used to define all unknown or uncertain variables. MCMC algorithms (e.g. Gibbs sampling) are used to sample from the conditional (posterior) probability distribution of each variable in succession, given all the other variables in the statistical model, in order to generate samples from the posterior distribution of the quantities of interest. The use of WinBUGS, however, is not a magic wand; it is a very powerful tool, but – as with any black box software – if you put rubbish in, you will get rubbish out. In particular, the estimated posterior distributions will be sensitive to the statistical model and prior distributions that the user defines. Care must be taken, therefore, when formulating the statistical relationships and prior distributions involved in the analysis.

For this study, prior distributions were assigned to all the uncertain parameters – $\mathbf{e} = [e_1, e_2, \ldots, e_P]$ and σ – and Bayesian inference was used to compare the regression model outputs against the empirical energy data ($\mathbf{E} = [E_1, E_2, \ldots, E_N]$). The result of the Bayesian inference is a set of posterior distributions (in the form of samples, which can be used to estimate probability density functions), which quantify the uncertainty for the energy intensities, **e**, and σ, and show the most probable values for these parameters based on the given prior knowledge and observed data.

The posterior distributions for the energy intensities ($\mathbf{e} = [e_1, e_2, \ldots, e_P]$) can then be used to provide a series of samples that act as 'virtual data' to calibrate the bottom-up engineering-based models of various building classes. In other words, the macro-level energy consumption data can be used to synthesize (or infer) 'observations' at the requisite resolution for calibrating a micro-level energy models; this calibration process is demonstrated in section 4.2. Additionally, the energy intensity estimates from the Bayesian regression analysis can be used as benchmark or reference values for the energy consumption of different domestic building classes. The advantage being that the benchmark values can be easily updated as new macro-level data becomes available annually. In either case, the use of probability distributions (i.e. the posteriors) to represent the energy intensity of a building class – instead of a single point estimate – allows the heterogeneity of dwellings within the building class – e.g. the variation in the physical characteristics of dwellings, their orientation, and the occupants – to be captured and taken into account.

2.3. *Case study*

A case study example was chosen to demonstrate the Bayesian regression method. For this study, domestic gas and electricity consumption data from 2008 at the LLSOA was used. The area of Salford in Greater Manchester, UK, containing a total of 144 LLSOAs, was chosen as the case study example. The gas and electricity consumption data was summed to derive the total annual energy consumption for each LLSOA[5] and outlying values were removed, giving a total of 137 LLSOA energy data points (i.e. $N = 137$). The gas and electricity consumption data for the case study LLSOAs in Salford are shown in Figures 1 and 2, respectively.

The building class for each dwelling was obtained from the results of an architectural survey which were recorded in geospatial mapping data,[6] which gives the structural type and construction period of each building, resulting in a total of 108 possible building classes. For the purpose of this study, similar dwelling types were merged into a single classification, whilst 10 infrequently occurring building classes – which in total accounted for only 0.6% of the dwelling floor area in Salford – were considered negligible and were discarded from the analysis.[7] This resulted in a total of 21 building classes (i.e. $P = 21$), consisting of four structural types (flats, terraced houses, semi-detached houses, and detached houses) and five construction periods (1870–1914, 1914–1945, 1945–1964, 1964–1979, 1979–present), with one additional class of modern cottage flats. Finer granularity for the construction periods would be desirable, in order to

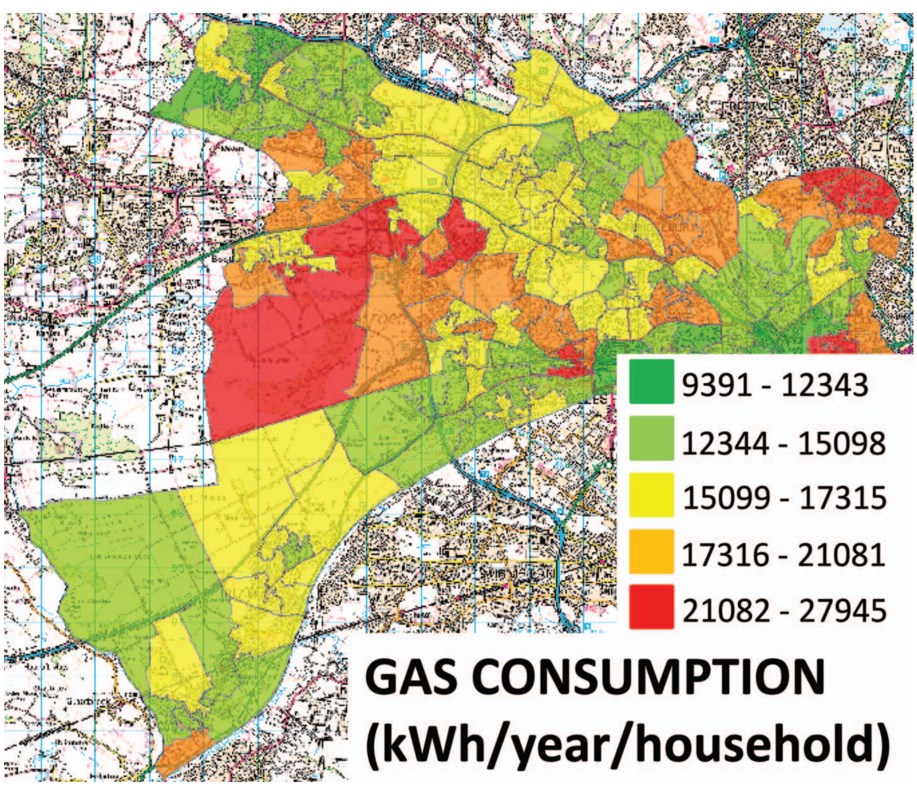

Figure 1. Gas consumption data for Salford, UK.

take into account changes to construction methods and energy standards that have been introduced over the last few decades; however, this was not possible, since the architectural survey data used to classify dwellings does not break the modern time period (1979–present) down into small sub-periods.

Geospatial mapping data also provide the footprint floor areas (in m^2) and building heights (in m) for all dwellings in the Salford area. An average floor height of 3.0 m was assumed to calculate the number of floors per dwelling, and this value was multiplied by the footprint floor area to obtain a value of the total floor area for each dwelling.[8] GIS software was used to calculate the total floor area for each building class in each LLSOA (x_{ij}), resulting in an $\mathbf{X}$ matrix with dimensions $[N, P]$.

Figure 3 shows an example of the geospatial mapping data in GIS software, with buildings colour coded by building class and a database that includes the following information for each building: the LLSOA; building height; building class; construction age; structural type; footprint area; estimated number of floors; and total floor area per dwelling.

Details of the 21 building classes are given in Table 1, including the structural type and the construction age for each class. In addition, Table 1 shows the relative dominance of each building class by giving the total floor area for each building class as a percentage of the total floor area for all dwellings in the Salford area. It can be seen that three building classes dominate the housing stock for this case study – pre-1914 terraced houses (building class 2), 1914–1945 semi-detached houses (building class 7), and 1945–1964 semi-detached houses (building class 11) – which between them account for 57.5% of the total floor area of dwellings in Salford.

Prior information on the energy intensity for each building class (e_j) was obtained from three sources:

(1) Predictions given by the Community Domestic Energy Model (CDEM) outlined in Firth *et al.* (2010), which gives estimates of the energy use by built form for the 2001 English housing stock, along with the average dwelling total floor area for each built form. CDEM is a bottom-up engineering-based housing stock model, based upon the BRE Domestic Energy Model (BREDEM) Kavgic *et al.* (2010).
(2) Domestic energy benchmarks from the Energy Savings Trust (EST),[9] which are also derived using BREDEM.
(3) Predictions given by an engineering-based building-level domestic energy model, developed by the authors of this paper, called SUSDEM

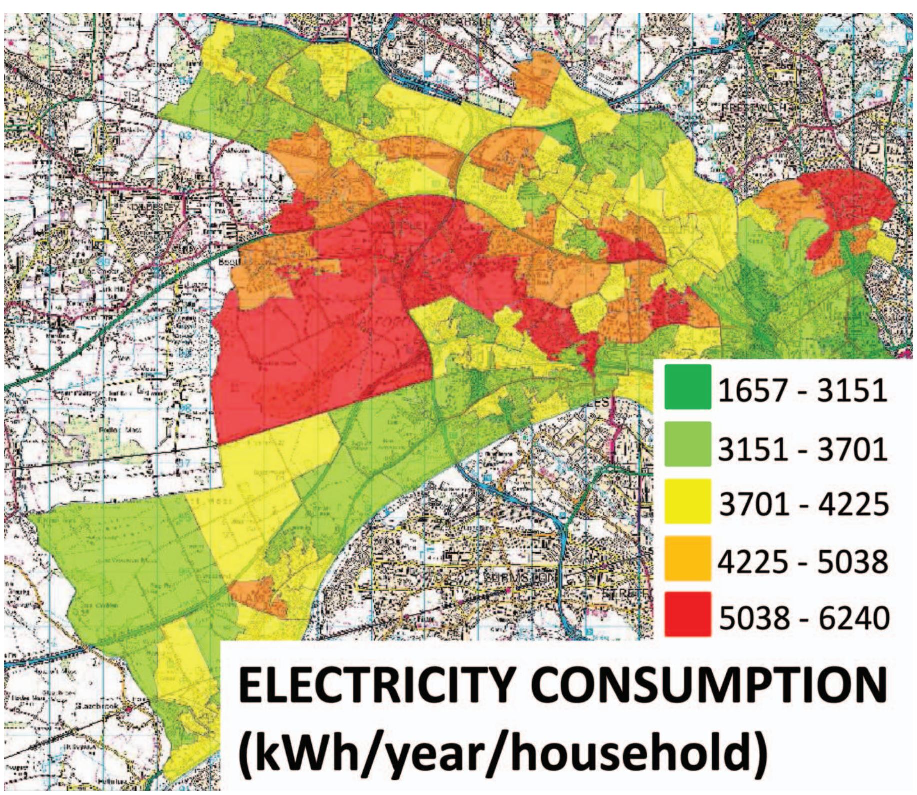

Figure 2. Electricity consumption data for Salford, UK.

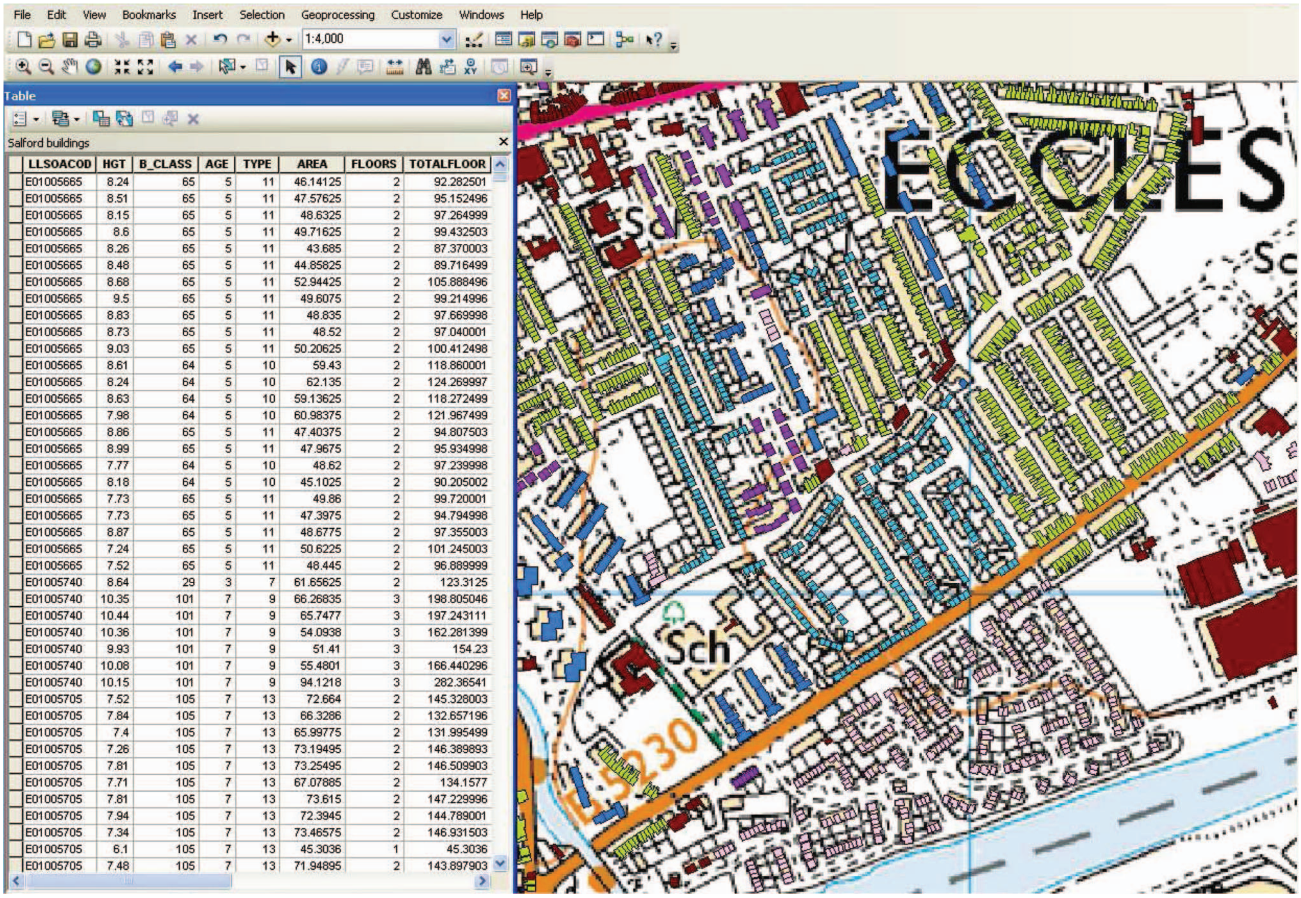

Figure 3. Example of geospatial mapping data in GIS software.

(stochastic urban-scale domestic energy model), which is outlined in Booth *et al.* (2012). SUSDEM is a quasi-steady state end-use energy model based on the assessment methodology provided by the CEN–ISO standards.[10] The CEN-ISO standards (e.g. ISO 13790) also form the basis of the BREDEM calculations for space heating.

The CDEM and EST sources give similar estimates for the energy intensity per building class, with an average energy intensity of approximately 260 kWh/m^2/year across all housing types, whilst the SUSDEM predictions gave estimates that were significantly lower, with an average of approximately 160 kWh/m^2/year. The similarity between the CDEM and EST sources is unsurprising, since these are both based on the same calculation procedure (i.e. BREDEM) and assumptions regarding inputs. The energy consumption calculations for SUSDEM are similar to those in BREDEM, but there are a few keys differences regarding the inputs to the model, which most likely explain the differences between the CDEM/EST sources and the SUSDEM predictions:

(1) The key difference concerns the assumptions regarding heating patterns and internal temperatures. As is demonstrated by the sensitivity analyses in both Firth *et al.* (2010) and Booth *et al.* (2012), the inputs relating to heating patterns and internal temperatures are key parameters in determining the predicted energy consumption. BREDEM-based estimates, such as those given by CDEM or the EST, make a certain set of assumptions regarding heating patterns and internal temperatures,[11] which may be true as a national average, but do not reflect the variation that exists between households and occupants. For example, one assumption used in BREDEM regarding heating preferences is that the living area will be heated to a temperature of 21°C during heating periods (i.e. when the heating is on); but this condition is unlikely to be true in many lower income households, as discussed in Milne and Boardman (2000) and Oreszczyn *et al.* (2006).

SUSDEM does not make the same assumptions regarding heating preferences. Instead,

Table 1. Case study building classes and prior information

Building class	Structural type	Construction age	% of total floor area, $(\sum_{i=1}^{N} x_{ij})/(\sum_{i=1}^{N}\sum_{j=1}^{p} x_{ij})$	Prior set 1: CDEM predictions and EST benchmarks: Mean energy intensity, $\bar{e}_j$ (kWh/m^2/year)	β parameters[12] α_1	α_2	γ parameters[13] a	b
1	Flat	1870–1914	0.4	260	41	59	50	5.2
2	Terraced	1870–1914	21.1	270	43	57	50	5.4
3	Semi	1870–1914	4.2	290	48	52	50	5.8
4	Detached	1870–1914	5.0	300	50	50	50	6.0
5	Flat	1914–1945	0.3	250	39	61	50	5.0
6	Terraced	1914–1945	0.2	260	41	59	50	5.2
7	Semi	1914–1945	25.9	280	46	54	50	5.6
8	Detached	1914–1945	2.0	290	48	52	50	5.8
9	Flat	1945–1964	1.8	240	37	63	50	4.8
10	Terraced	1945–1964	1.4	250	39	61	50	5.0
11	Semi	1945–1964	10.5	270	43	57	50	5.4
12	Detached	1945–1964	0.4	280	46	54	50	5.6
13	Flat	1964–1979	3.2	230	34	66	50	4.6
14	Terraced	1964–1979	5.4	240	37	63	50	4.8
15	Semi	1964–1979	3.9	260	41	59	50	5.2
16	Detached	1964–1979	3.0	270	43	57	50	5.4
17	Flat	1979–2011	1.0	220	32	68	50	4.4
18	Terraced	1979–2011	1.4	230	34	66	50	4.6
19	Cottage flat	1979–2011	4.3	220	32	68	50	4.4
20	Semi	1979–2011	1.4	250	39	61	50	5.0
21	Detached	1979–2011	3.4	260	41	59	50	5.2

Building class	Structural type	Construction age	% of total floor area, $(\sum_{i=1}^{N} x_{ij})/(\sum_{i=1}^{N}\sum_{j=1}^{p} x_{ij})$	Prior set 2: SUSDEM predictions: Mean energy intensity, $\bar{e}_j$ (kWh/m^2/year)	β parameters[12] α_1	α_2	γ parameters[13] a	b
1	Flat	1870–1914	0.4	160	19	81	50	3.2
2	Terraced	1870–1914	21.1	170	21	79	50	3.4
3	Semi	1870–1914	4.2	190	26	74	50	3.8
4	Detached	1870–1914	5.0	200	28	72	50	4.0
5	Flat	1914–1945	0.3	150	17	83	50	3.0
6	Terraced	1914–1945	0.2	160	19	81	50	3.2
7	Semi	1914–1945	25.9	180	23	77	50	3.6
8	Detached	1914–1945	2.0	190	26	74	50	3.8
9	Flat	1945–1964	1.8	140	14	86	50	2.8
10	Terraced	1945–1964	1.4	150	17	83	50	3.0
11	Semi	1945–1964	10.5	170	21	79	50	3.4
12	Detached	1945–1964	0.4	180	23	77	50	3.6
13	Flat	1964–1979	3.2	130	12	88	50	2.6
14	Terraced	1964–1979	5.4	140	14	86	50	2.8
15	Semi	1964–1979	3.9	160	19	81	50	3.2
16	Detached	1964–1979	3.0	170	21	79	50	3.4
17	Flat	1979–2011	1.0	120	10	90	50	2.4
18	Terraced	1979–2011	1.4	130	12	88	50	2.6
19	Cottage flat	1979–2011	4.3	120	10	90	50	2.4
20	Semi	1979–2011	1.4	150	17	83	50	3.0
21	Detached	1979–2011	3.4	160	19	81	50	3.2

the parameters defining heating patterns and internal temperatures are calibrated on a case-by-case basis using a Bayesian calibration method and empirical energy consumption data, as detailed in Booth *et al.* (2012). The SUSDEM predictions used in this study are based on a model that was calibrated using energy consumption measurements from a group of low-income flats in Salford, where the average internal temperatures were observed to be below 21°C in the living areas.

(2) Assumptions regarding other input parameters also differ between the CDEM/EST estimates and the SUSDEM predictions; for example,

> CDEM and the EST use national statistics (e.g. from the English House Condition Survey) for key input parameters relating to the building envelope (e.g. wall U-value) and the building systems (e.g. boiler efficiency). SUSDEM, however, again calibrates these uncertain input parameters on a case-by-case basis, similarly to the calibration of the parameters for heating conditions, as explained above.

It is these differences relating to input assumptions that explain the differences between the CDEM/EST estimates and the SUSDEM predictions. It is not possible to know *a priori* which source of information best reflects the case study housing stock. Whilst the SUSDEM predictions were generated from a calibrated model, this calibration procedure was conducted using a sample of low-income flats, which may not be representative of the larger Salford housing stock. In this respect, the CDEM/EST estimates – which use national statistics as inputs to calculate energy consumptions – may be more representative of the case study housing stock.

For the reasons outlined above, these three sources of information were not combined into one set of priors. Instead two separate sets of priors were used – one set based upon a combination of the information given by the CDEM predictions and the EST benchmarks, and another set based upon the predictions from SUSDEM.

In addition, for both sets of prior information, two separate distributions were tested – one set of β (beta) distributions and another set of γ (gamma) distributions – to examine the effect of different priors on the resultant set of posteriors. The main difference in the β and γ distributions is in their bounding – β distributions[12] are bound between 0 and 1, therefore all energy consumption and energy intensity values have to be scaled, whilst γ distributions[13] are only given a lower bound of zero, with no upper bound, therefore no scaling is required. This gave a total of four sets of priors – two sets of β distributions for the CDEM/EST sources and for SUSDEM predictions, and two sets of γ distributions.

Prior information was also specified for σ, the unknown standard deviation shown in Equation (3a). The prior for σ was the same in all cases, and was specified using a γ distribution[13] as follows:

$$\tau \sim \gamma(a = 10, b = 0.01) \tag{5}$$

where $\tau = \frac{1}{\sigma^2}$.

This equates to a low expected value for the prior of σ of approximately 3.2 kWh/m^2/year with a fairly low spread of possible values (standard deviation of approximately 1.0 kWh/m^2/year), as shown in Figure 4. Given

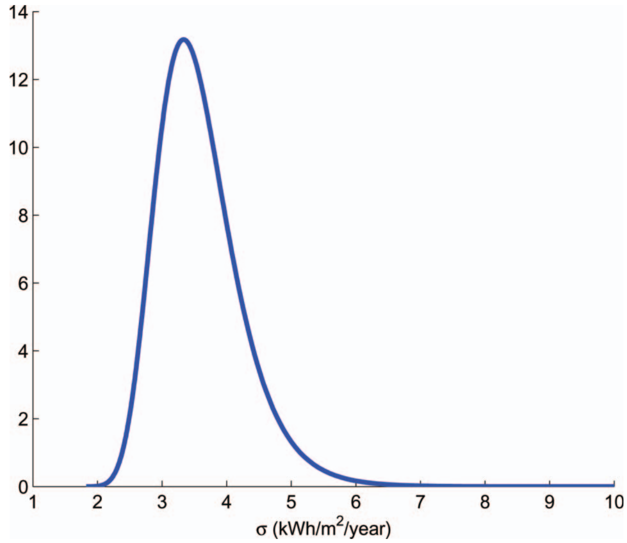

Figure 4. Prior for σ.

that the empirical energy data is assumed to be a 'known' variable, it is reasonable to assume a low expected value for the prior of σ. However, the use of a γ distribution, as specified in Equation (5) and shown in Figure 4, provides a long upper tail for the prior of σ, thus allowing for the possibility of a far greater value of σ.

Information on the priors for all 21 building classes, including the mean energy intensity of the prior distributions, and the values of α_1 and α_2 for the β distributions and a and b for the γ distributions, is shown in Table 1. The results of the Bayesian regression analysis for this case study are shown in Figure 3.

3. Results and discussion

3.1. Energy intensity posteriors

The results of the Bayesian regression can be seen in Figure 5, which shows the posterior and prior distributions for the energy intensities of all building classes using the β distribution priors (as outlined in Table 1). β_1 (shown in red) is the set of energy intensities for the CDEM predictions and EST benchmarks, whilst β_2 (shown in blue) is the set for the predictions from SUSDEM. Prior distributions are given by the solid curves, whilst posterior distributions are shown by the dashed curves.

The key feature to note in Figure 5 is that the greatest differences between the prior and posterior distributions of energy intensity are seen for the most dominant building classes – i.e. the building classes with the greatest total floor areas, which make up the majority of the stock, and which appear in a larger number of LLSOAs. The dominant building classes are: pre-1914 terraced houses, e_2; 1914–1945 semi-detached houses, e_7; and 1945–1964 semi-detached

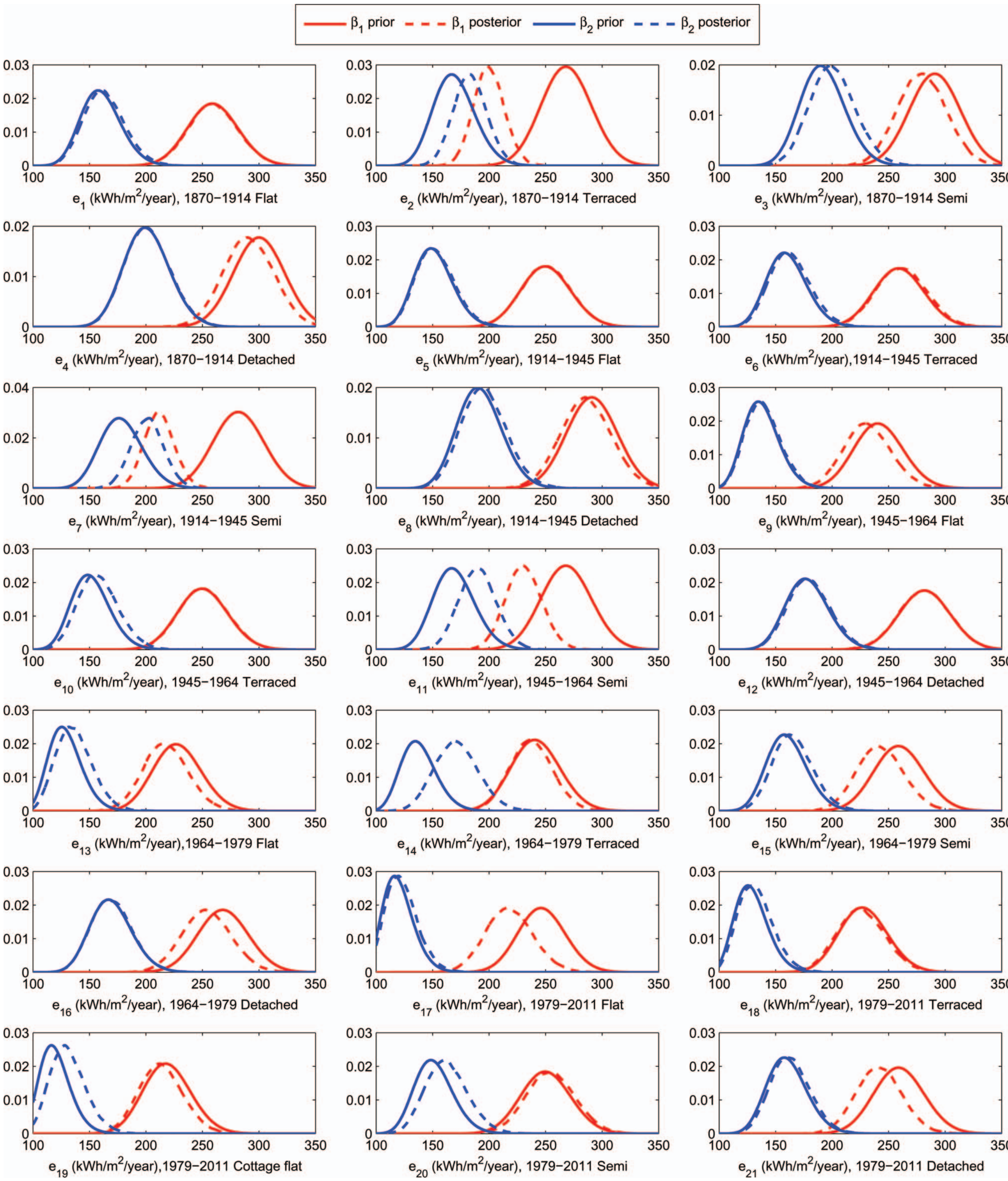

Figure 5. Prior and posterior distributions of energy intensity using β priors.

houses, e_{11}. For these dominant building classes, the two derived posteriors converge towards a common result, with similar – but not identical – posteriors from the Bayesian regression analyses, despite the difference in the prior distributions. For example, it can be seen that the posteriors of energy intensity for the most dominant building class – 1914–1945 semi-detached houses, e_7, which makes up of 25% of the total floor area for dwellings in Salford – converge very strongly, producing similar posterior distributions. However, it can also be seen that the convergence of the posteriors for e_{11} – which is the less dominant than the e_2 and e_7 building classes, with 10.5% of the total case study floor area – is quite weak, with quite a large difference still seen between the two posterior distributions.

In addition, the less dominant building classes show little difference between the priors and the posteriors. This demonstrates a limitation of the Bayesian regression method regarding its ability to produce accurate energy intensity estimates for the less frequent building classes.

These outcomes indicate that the inference is strengthened by the increased availability of information. The dominant building classes are found in a greater number of LLSOAs; therefore there is more information for the Bayesian regression to utilize when estimating the energy intensities of these building classes. The stronger and more robust the available empirical energy data is, the weaker the effect of the subjective information of the prior distribution is on the resultant posterior distribution. Conversely, it is also the case that the stronger the given prior information, the weaker the effect of additional empirical data will be.

It can also be seen that for the dominant building classes, where information is strongest, the distributions for the energy intensity have narrowed, showing that the uncertainty regarding the average value of the energy intensity has been reduced for these building classes. The wider posterior distributions for the less dominant building classes indicate that there is a greater uncertainty about the energy intensity estimates for these classes, again showing the limitations of the Bayesian regression method (in its most simplified form) regarding its ability to generate accurate estimates for the less frequent building classes. This weakness of the Bayesian regression method is addressed in section 3.3.

The analysis shows that the average energy intensity for each building class is likely to be somewhere in between the expected values of prior set 1 and prior set 2 (as shown in Table 1). This is clearly shown in Figure 5 by the convergence towards the middle-ground of the energy intensity distributions for the dominant building classes (such as e_2, e_7, and e_{11}), and also by the movement of the distributions for the less dominant classes, such as e_{17} and e_{21}, where β_1 is shifted to the left, and e_{14} and e_{19}, where β_2 is shifted to the right.

Results for the Bayesian regression using the γ distribution are similar and are not shown in their entirety to avoid duplication; however, an example is shown in Figure 6 of the prior and posterior distributions of energy intensity for 1914–1964 semi-detached houses (e_7) using γ priors. A comparison between the posteriors in Figure 6 (shown by the blue and red dashed curves) and the posteriors for the same building class, e_7, in Figure 5 shows that similar posterior distributions are derived from the Bayesian regression analyses despite the use of two different parametric families of curves – i.e. β distributions and γ distributions – used to define the prior distributions (shown by the blue and red solid curves). The comparison between the use of β and γ distributions is shown for prior set 1 for the same building class, e_7, in Figure 7. Figure 7 demonstrates this with more clarity by showing the results of the Bayesian regression analyses for prior set 1 using a β distribution and a γ distribution on the same graph. The solid blue and solid red curves indicate the prior distributions using a β distribution and a γ distribution, respectively, whilst the dashed blue and dashed red curves are the associated posteriors.

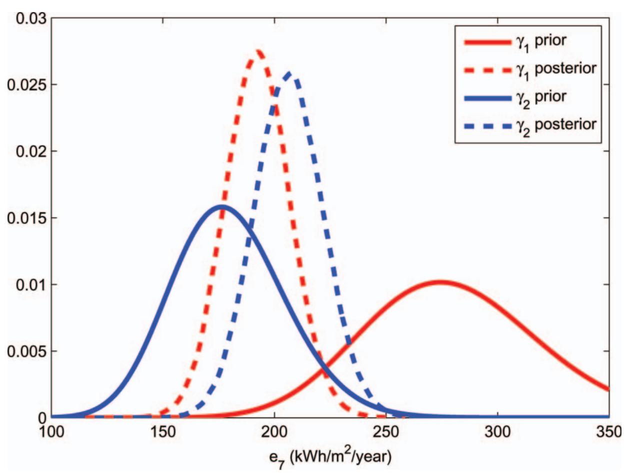

Figure 6. Prior and posterior distributions of energy intensity for 1914–1964 semi-detached houses using γ priors.

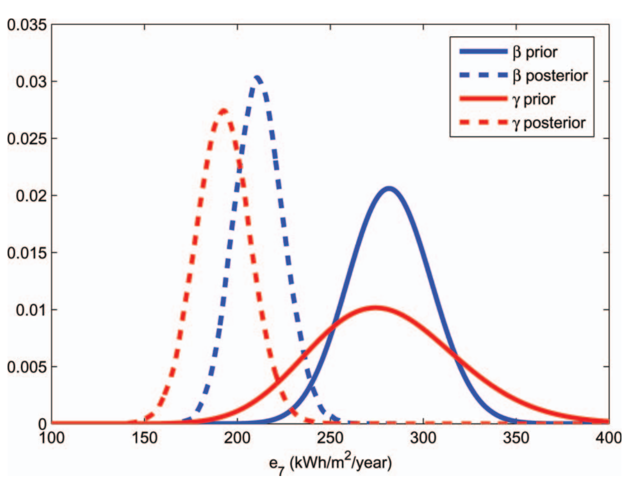

Figure 7. Comparison of priors and posteriors for e_7 using β and γ distributions from prior set 1.

3.2. Combining priors

Prior information represents expert opinion, and is quantified in the form of prior distributions through a process of elicitation (O'Hagan and Oakley 2004, O'Hagan 2006). These prior beliefs can be obtained from a variety of sources, including previous studies and literature review, rules-of-thumb and guidelines, academic and professional opinion, and data from a small subset sample of the population. In many cases, prior information is available from multiple sources. In such cases, elicitation can be used to quantify the multiple sources of information into a single prior distribution.

The strength of any single prior distribution will depend on whether there is an agreement between the various sources of prior information. For example, if the different sources of prior information are conflicting, then a weak prior distribution (such as a uniform prior with a wide range) will be specified and the posterior distribution will be more strongly affected by the empirical data. Conversely, if the different sources of prior information are in close agreement, then a very strong prior can be defined and the posterior distribution will only be weakly affected by the empirical data.

As outlined in section 2.3, the prior information in this case study came primarily from three sources: predictions from CDEM, as given in Firth *et al.* (2010); energy benchmarks from the EST; and predictions from SUSDEM, as described in Booth *et al.* (2012). Two of these sources – the CDEM predictions and the EST benchmarks – show close agreement and have been amalgamated into a single prior distribution (prior set 1, shown as a solid red line in Figures 5 and 6); the other source – the SUSDEM predictions – show a disparity against the other two sources, and have therefore been represented as a separate prior distribution (prior set 2, shown as a solid blue line in Figures 5 and 6).

In essence, therefore, the results presented in section 3.1 and Figure 5 are actually the results of two separate Bayesian regression analyses – with two separate sets of priors – that have been overlaid onto a single set of graphs. One of the great strengths of Bayesian inference and the WinBUGS software, however, is that it enables multiple sources of information – both in terms of prior knowledge and empirical data – to be combined and utilized into a single analysis without having to specify a single prior distribution to represent the various sources of information. For example, in the case study demonstrated above, the two sets of prior distributions can be specified separately and then combined into a single Bayesian regression analysis, producing a single set of posterior distributions.

The process of combining prior sets into a single Bayesian regression analysis was applied to the case study. This is conducted in WinBUGS by using an algorithm that instructs the software to randomly pick from either one set of priors or the other when conducting the Bayesian analysis. An initial weighting is assigned for each prior set, defined by a probability, which specifies the chance of picking from each prior set during the Bayesian analysis. This expresses the initial confidence that the analyst has in each prior set; for example, in this case, a 50% probability was assigned in terms of the chance of selecting either prior set, indicating that both prior sets were initially thought to be equally valid. This initial (prior) probability is updated during the Bayesian analysis on the basis of how well each prior set fits the empirical data – i.e. the energy consumption data and the measured floor areas – in order to generate the final (combined) posterior distribution for the quantities of interest – i.e. the energy intensities for the building classes. The results of this analysis are shown in Figure 8 for the combination of the two sets of γ priors (as specified in Table 1).

There are two notable features regarding the posterior distributions (shown as green dashed lines) in Figure 8. Firstly, it can be seen that for almost all building classes, the posterior distribution of the energy intensity is biased towards prior set 2 – i.e. towards the priors given by the SUSDEM predictions. This indicates that the energy intensities (**e**) specified by prior set 2 provide a better fit between the energy consumption data (**E**) and the measured floor areas (**X**) for the regression relationship given in Equation (3a). This, in turn, indicates that the actual energy intensity for most building classes will likely be towards the lower end of the range covered by the prior knowledge.

Secondly, it can once again be observed that the uncertainty regarding the energy intensity – given by the spread of the posterior distribution – is reduced most greatly for the dominant building classes, such as the pre-1914 terraced houses (e_2) and 1914–1945 semi-detached houses (e_7). Nevertheless, as with the results in sections 3.1 and 5, there is still a large range for these posteriors of approximately 100 kWh/m^2/year. Meanwhile, for the less dominant building classes where the evidence is weaker, it can be seen that the posterior distributions have very wide spreads, indicating that there is still a great deal of uncertainty over the energy intensity for these building classes. In fact, for some building classes, such as e_1, e_{10}, and e_{20}, there is an almost bi-modal posterior distribution with two maxima, which indicates that the posterior has been only weakly affected by the empirical data and is mainly influenced by the mixture of the two prior sets.

The combined analysis provides greater confidence that the energy intensities are towards the lower end of

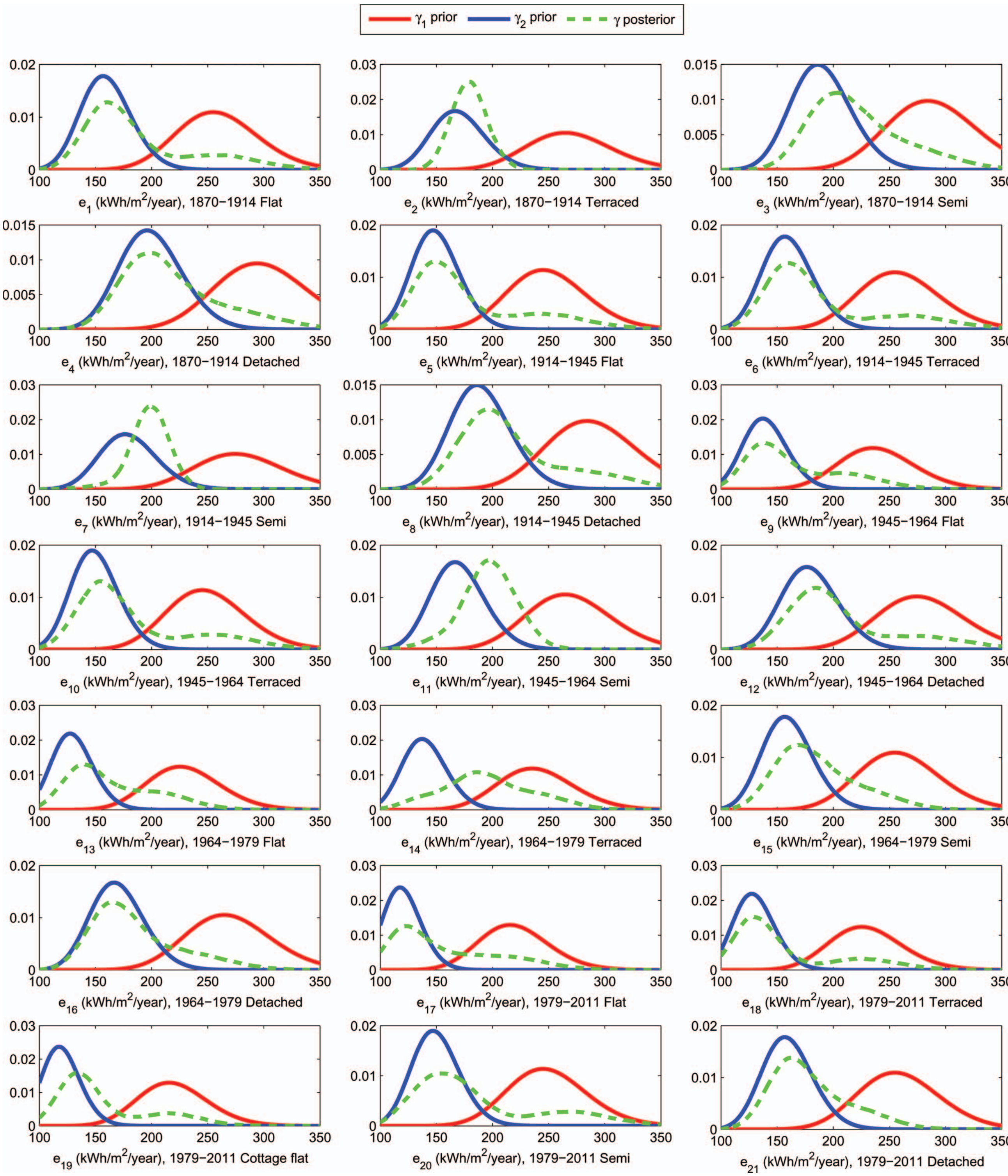

Figure 8. Prior and posterior distributions of energy intensity using a combination of two sets of γ priors.

the spectrum, whilst also producing a unimodal posterior distribution of energy intensity from multiple sources of prior information for the more dominant building classes.

The strength of Bayesian inference and its ability to combine multiple sources of information goes far beyond what has been outlined and examined above. Not only can multiple sources of prior information be combined into a single analysis, but multiple sources of observed data can also be included. The different levels of confidence associated with the different sources of empirical data can be taken into account by weighting the various sources of evidence according to their reliability, as explained in Spiegelhalter and Best (2003). For example, if a more reliable set of energy consumption data was made available in the context of the case study above – i.e. if the process used to collect the LLSOA gas and electricity consumption data was improved and applied to collect an additional year's worth of energy data – then this could be added to the existing energy consumption data in the Bayesian regression analysis, but given a higher weighting to signify its greater reliability. Meanwhile, sources of observed data at different scales can also be integrated into a single analysis using Bayesian inference, as explained in Jackson *et al*. (2006). In the context of this case study, this would mean using energy data that was reliably collected at the building or district scale – i.e. for a small group of similar dwellings from the same building class (e.g. 1914–1945 semi-detached houses) – to be combined in the same analysis with the DECC energy consumption data that is collected at an urban or national scale.

Finally, it is also possible to use different sets of covariates (i.e. independent variables) within the same Bayesian regression analysis, as explained in Jackson *et al*. (2009). For example, if there was another source of information available that could provide an additional set of estimates for the floor areas of the different building class in the case study above – i.e. another set of data for **X** – then this additional source could be integrated into the analysis to estimate the energy intensities (the quantity of interest) with more certainty.

It can be seen, therefore, that the Bayesian regression method does not have to be seen as a single, isolated, static analysis, but can be thought of as an ongoing process in which uncertainty is continually being reduced as additional, more reliable sources of information become available and are integrated into the analysis.

3.3. Uncertainty in measured data

Any linear regression model, such as that shown in Equation (1), will consist of: a set of unknown parameters that are of interest and are to be determined; a set of independent variables, which are assumed to be known from empirical data; and a set of dependent variables, which are also assumed to be known from empirical data. In the case of Equation (1), the energy intensities (**e**) are the unknown variables, the percentage floor areas (**X**) are the independent variables, and the empirical energy consumption data (**E**) is the dependent variables.

In practice, however, the independent and dependent variables are never completely 'known'; both sets of data will contain uncertainties. As mentioned previously in section 2.2, a Bayesian method allows for uncertainty over the dependent variables to be taken into account and quantified. For this study, this is achieved by representing the energy consumption data, E_i, as being drawn from a probabilistic distribution with mean value μ_i and standard deviation σ, as shown in Equation (3a).

Figure 9 shows the posterior for σ (i.e. the standard deviation of the energy consumption) that results from the Bayesian regression analysis presented in section 3.2 and Figure 8 (using a combination of prior sets for the energy intensities). This analysis used the prior distribution for σ specified in Equation (5) and shown in Figure 4.

By comparing Figure 9 against Figure 4, it can be seen that, having specified a prior for σ with an expected value of approximately 3.2 kWh/m^2/year and a standard deviation of approximately 1.0 kWh/m^2/year, the posterior distribution for σ is in the range of 60–80 kWh/m^2/year, with an expected value for σ of 69 kWh/m^2/year and a standard deviation of 4.6 kWh/m^2/year. This gives a ratio of the mean standard deviation (i.e. the mean of σ) to the mean of the energy data (i.e. $\sum_{i=1}^{N} E/N$) – given overall by $(N\bar{\sigma})/(\sum_{i=1}^{N} E)$ – of approximately 1/3, indicating that there is either a great deal of inaccuracy over the empirical energy consumption data or that the regression is unable to reconcile the given combinations of **E** and **X**.

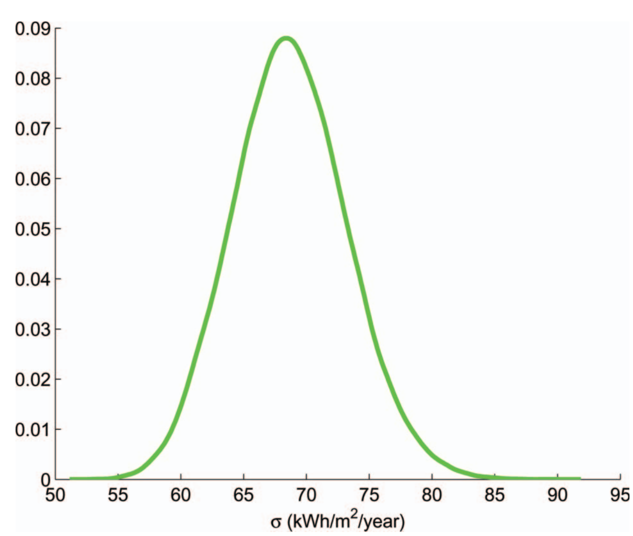

Figure 9. Posterior for σ using a combination of γ priors for the energy intensities.

The most likely explanation is a combination of both. It is known that there is a great deal of inaccuracy over the energy consumption data, arising from the method by which it is collected. This method is outlined for domestic energy consumption data in DECC (2010, 2011), whilst a similar process is used in Bruhns and Wyatt (2011) for non-domestic energy consumption data. DECC themselves classify the method for collecting this data as 'developmental' and classify the statistics produced as 'experimental' (DECC 2011).

These potential inaccuracies in the energy consumption data arise from the data collection methodology in a number of ways. Firstly, the gas and electricity consumption figures given in the DECC data set are obtained from the agents of gas and electricity suppliers, who generate 'annualized' consumption data (which is used to generate gas and electricity bills by the suppliers), rather than using actual complete measured annual data. Although meter readings are taken regularly by gas and electricity agents, these meter readings rarely cover an exact and complete annual period (i.e. a complete calendar year). Instead, annualized consumption data for a given calendar year is estimated on the basis of either an 'annualized advance' (AA) or an 'estimated annual consumption' (EAC), as explained in DECC (2010). The AA estimates annual consumption by extrapolating the consumption recorded between two meter readings, whilst the EAC is used when only a single meter reading is available. In this latter case, the EAC is estimated by comparing the single meter reading against historical profile information for the given property's meter. The energy consumption data, therefore, is not based on actual complete measured annual data for the calendar year in question. In addition, the gas consumption data is 'weather corrected', with the consumption data adjusted by comparing the heating degree days for the year in question against a 17-year average. This means that the figures given for gas consumption do not necessarily correspond to the actual gas consumption associated with the weather for the year that the data was collected.

Secondly, the method used to collect energy data requires the data collectors to classify meters as belonging to either domestic or non-domestic properties. This is more reliably done for electricity consumption readings, since domestic meters always operate on a non-half-hourly basis, whilst large industrial and commercial customers use half-hourly meters. However, there is still uncertainty about the domestic/non-domestic classification of electricity meters, since many small- and medium-sized businesses will be on similar non-half-hourly meters to domestic users. The DECC method, therefore, chooses to classify all meter readings with an annualized electricity consumption of over 100 MWh as being non-domestic electricity consumption. For gas meter readings, the process is less reliable, since the classification of gas meters as domestic/non-domestic is not as clearly defined. As a result, the DECC method uses the gas industry standard of classifying all annualized gas consumption of over 73.2 MWh as being non-domestic.

This method for classifying gas and electricity meters as either domestic or non-domestic clearly leads to inaccuracies in the energy consumption data. For example, there may be cases of buildings with a single gas or electricity meter that is shared between domestic and non-domestic users, e.g. a small shop with a flat above. In other cases, it is possible that meters belonging to businesses with a low energy consumption will be classified as domestic, whilst meters in residences with a high energy consumption may be classified as non-domestic.

Finally, it is possible that the energy data collection method does not include certain buildings, and therefore that some energy consumption is missing from the figures, which would result in the data providing an underestimate of the actual energy consumption.

In addition to uncertainties in the empirical energy data, there are also many sources of uncertainty in the total floor areas calculated from the geographic information and mapping data; this uncertainty, however, has so far not been taken into account in the Bayesian regression method described above, where x_{ij} are considered to be 'known' input variables.

For example, the total floor areas were calculated by multiplying the footprint area from the mapping data by the number of floors, which were in turn estimated from the height of the building. Not only is there uncertainty over the measured height of the building, but there is also uncertainty over the calculation to convert this height into a number of floors. In addition, there is uncertainty about the footprint area and whether this is the correct area to use, since the actual internal dwelling floor area will be smaller. Finally, there is also the possibility that the mapping data does not include certain properties – or includes too many properties – due to an incorrect classification of buildings as either domestic or non-domestic.

These sources of uncertainty that have been described above show that the input variables – i.e. the energy consumption data and the floor areas – are far from 'known' and that uncertainty in the measured data should be accounted for and quantified in order to obtain more accurate estimates for the unknown parameters of interest, i.e. the energy intensities, e_j.

In order to examine the magnitude and the effect of these uncertainties, the Bayesian regression method outlined in sections 2.2 and 3.2 was developed to include uncertainty in measured variables using a Bayesian 'errors-in-variables' model, as outlined in Dellaportas and Stephens (1995). In the error-in-variables model for this study, the x_{ij} term in Equation (3b) is now treated as an uncertain variable, and is stochastically specified, such that:

$$x_{ij} \sim N(\beta_j z_{ij}, \phi) \tag{6}$$

In Equation (6), z_{ij} is the measured floor area for each building class and each LLSOA, whilst x_{ij} is the 'true' floor area. β_j is an 'area correction factor', which is dependent on the building class. The assumptions behind the β_j term are that similar dwellings will have a consistent error between the measured floor area (z_{ij}) and the actual floor area (x_{ij}), and that this error will be a constant factor – i.e. the actual floor area will be a multiple of the measured floor area. A weak uniform prior was set for β_j of: $\beta_j \sim U(0.5, 1.25)$. The bias towards a value of $\beta_j < 1$ was made on the assumption that the measured areas are larger than the actual areas, since the actual internal dwelling floor area will be smaller than the footprint area measured from the mapping data. In addition, a weak prior was given for ϕ, the standard deviation of the actual floor area around $\beta_j z_{ij}$. The same prior was set for σ as was used in sections 3.1 and 3.2.

This 'error-in-variables' method produces posterior distributions not only for the energy intensities, but for the area correction factors (β_j) also, which allows for uncertainties in the mapping data to be taken into account.[14]

The overall framework for the Bayesian regression analysis with errors-in-variables is shown diagrammatically in the directed acyclic graph (DAG) in Figure 10. DAGs are used in Bayesian inference to express the structure of the model and to show the relationship between the different known and unknown quantities. Each parameter (both known and unknown) is displayed as an elliptical node in the DAG, with relationships between nodes shown by arrows. The direction of the arrow indicates the direction of the dependence, with solid arrows showing a stochastic dependence and hollow arrows indicating a deterministic relationship. Finally, dotted boxes – known as 'plates' – indicate repeated (i.e. looped) sections of the model. See Lunn *et al.* (2000) for a more detailed description of DAGs.

In addition to the inclusion of the error-in-variables, a different method was used for combining the two sets of energy intensity priors. In this case, the two sets of energy intensities (as specified in Table 1) were treated as data that were drawn from a common 'true' set of energy intensities, e_j, with some sort of

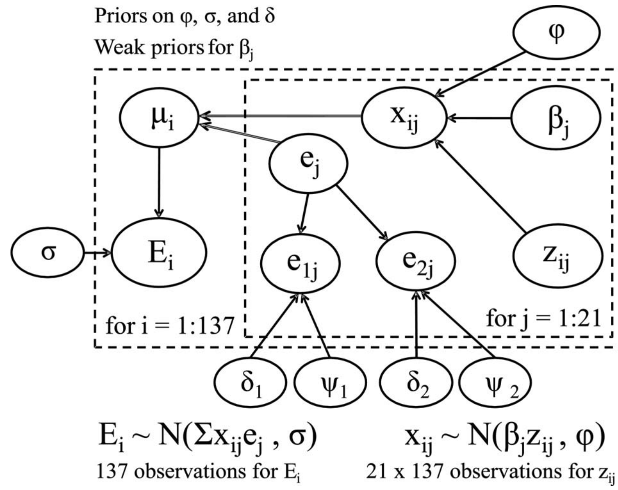

Figure 10. Directed acyclic graph for the Bayesian regression with errors-in-variables model.

constant bias (δ_1 and δ_2) for each set of energy intensities (e_{1j} and e_{2j}) and an associated standard deviation (ψ_1 and ψ_2):

$$e_{1j} \sim N(e_j + \delta_1, \psi_1) \tag{7a}$$

$$e_{2j} \sim N(e_j + \delta_2, \psi_2). \tag{7b}$$

A very weak uniform prior was set for the 'true' set of energy intensities, e_j, with normally distributed priors specified for the energy intensity biases, δ_1 and δ_2, around a mean value of zero – i.e. an initial assumption that there should be no constant bias. e_{1j}, e_{2j}, ψ_1, and ψ_2 were all specified based on the means and standard deviations of the prior distributions shown in Figure 8.

The resultant posteriors for the energy intensities (e_j) that are derived from the BRA with the error-in-variables are shown in Figure 11 (along with the two sets of prior distributions, as specified in Table 1). In addition, the mean of the energy intensity posteriors are shown for each in building class in Table 2, along with the mean of the posteriors for the area correction factors, β_j.

The most notable differences between the posteriors for the energy intensities shown in Figure 11 and those seen previously in Figure 8 are the more pronounced shift between the prior and posterior for the less dominant building classes – e.g. e_1, e_{10}, and e_{20}, along with the reduced spread for these posteriors. These differences are the result of the constant bias terms, δ_1 and δ_2, as specified in Equation (7) and Figure 10, which allow for information to be learnt about the energy intensities of all building classes, even for those classes that appear less frequently in the housing stock. In addition, it can be seen that the

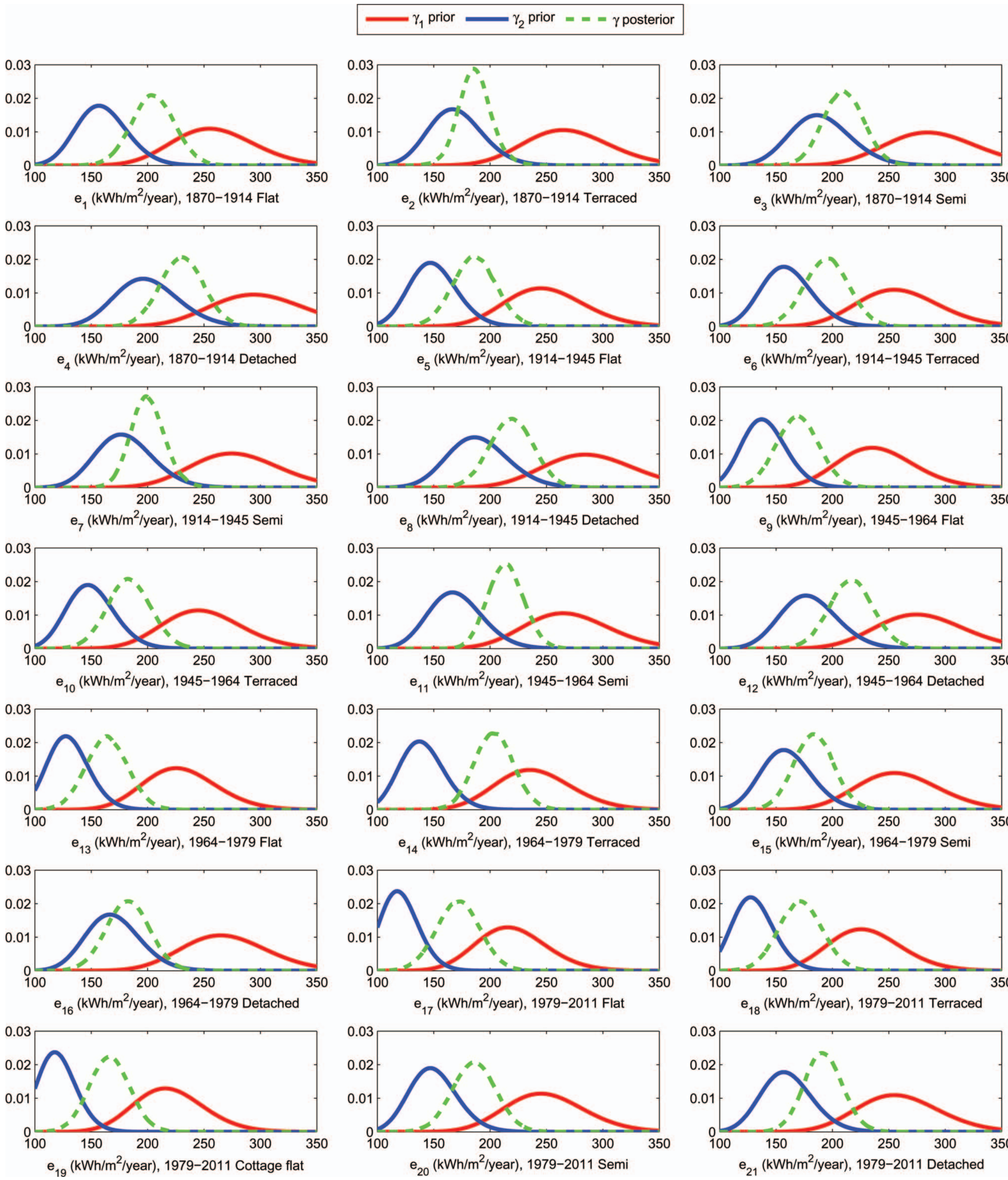

Figure 11. Prior and posterior distributions of energy intensity using a combination of two sets of γ priors for the Bayesian regression with error-in-variables model.

spread of the energy intensity posterior is narrower for the dominant building classes – e.g. e_2 and e_7 – indicating that there is greater certainty about the real value of energy intensity for these classes.

It can be seen in Table 2 that, despite the weak priors given for β_j, the mean estimates for the area correction factors are all in the range of 0.77–0.93 – i.e. that the 'true' floor areas are consistently smaller than

Table 2. Estimate energy intensities and area correction factors from the Bayesian regression with error-in-variables model.

Building class	Structural type	Construction age	Energy intensities e_j mean of posteriors (kWh/m^2/year)	Area correction factors, β_j mean of posteriors
1	Flat	1870–1914	205	0.87
2	Terraced	1870–1914	182	0.77
3	Semi	1870–1914	205	0.78
4	Detached	1870–1914	231	0.82
5	Flat	1914–1945	185	0.83
6	Terraced	1914–1945	195	0.83
7	Semi	1914–1945	192	0.77
8	Detached	1914–1945	219	0.81
9	Flat	1945–1964	177	0.84
10	Terraced	1945–1964	183	0.82
11	Semi	1945–1964	207	0.82
12	Detached	1945–1964	216	0.84
13	Flat	1964–1979	162	0.81
14	Terraced	1964–1979	199	0.93
15	Semi	1964–1979	179	0.79
16	Detached	1964–1979	187	0.78
17	Flat	1979–2011	171	0.88
18	Terraced	1979–2011	173	0.86
19	Cottage flat	1979–2011	166	0.87
20	Semi	1979–2011	189	0.85
21	Detached	1979–2011	205	0.85

the measured floor areas. This is to be expected, since the measured floor areas are estimated based on the footprint of the buildings, taken from the mapping data, which is larger than the actual internal dwelling area.

Table 3 shows a comparison between the energy intensity estimates from the Bayesian regression with error-in-variables against those given in Cheng and Steemers (2011), calculated for the north-west region of England (where Salford is located) using the Domestic Energy and Carbon Model (DECM), a bottom-up BREDEM-based housing stock model. Figures for both the BRA and DECM are shown by structural type and are given as mean estimates across all dwelling construction ages.

It can be seen from Table 3 that the BRA estimates are, on average, lower than the DECM estimates (with the exception of the detached structural type, where the pattern is reversed). There are a number of possible factors behind this difference. Importantly, it is very likely that the two sets of estimates are not directly comparable for several reasons. Firstly, it is noted in Cheng and Steemers (2011) that the DECM model over-estimates gas and electricity consumption when compared to the aggregated DECC energy consumption figures for the whole of Englandˆ. The main reason given by Cheng and Steemers (2011) for this disparity is the difference in the climate data used: the DECC energy consumption data is weather corrected using a 17-year average (1988–2004), whilst the DECM model uses a 30-year average (1961–1990); meanwhile the average winter temperatures for these two periods

Table 3. Comparison of energy intensity estimates from the Bayesian regression against results from Cheng and Steemers (2011).

Structural type	Energy intensity estimates (kWh/m^2/year): Mean of posterior	Energy intensity estimates (kWh/m^2/year): Chang and Steemers (2011), NW regional average
Flat	180	208
Terraced	186	235
Semi-detached	194	241
Detached	211	184

are 4.7°C and 3.7°C, respectively, leading to a lower heating demand for the former and a higher heating demand for the latter. The Bayesian regression uses this same empirical energy data from DECC, and is therefore also likely to lead to lower estimates for the energy intensities.

Finally, the DECC data used for the Bayesian regression was for Salford, UK, whilst the DECM estimates are an average for the entire north-west region of England; there are, therefore, possible socio-economic factors behind the difference in the energy estimates seen in Table 3. This factor – that of socio-economic differences – raises the important point that residential energy consumption is not merely a function of the structural type and construction age of the dwelling, but is also heavily influenced by internal activity and occupant behaviour, which vary greatly even for physically similar dwellings. Socio-economic factors that vary from region to region – and from dwelling to dwelling – can therefore play an important role in determining the energy consumption of a housing stock or of an individual dwelling; for example, household income may influence the internal heating set-point temperature preferences of occupants and their use of electrical appliances, with the result of wealthier households consuming a greater amount of gas and electricity.

4. Validation and future application

4.1. Validation of the Bayesian regression analysis

In order to use the energy intensity estimates generated (as discussed in section 3.3) for any further application, the Bayesian regression (with error-in-variables) method must be validated. For this study, a Monte Carlo cross validation method was used. Cross validation, as discussed in Arlot (2010) and Witten *et al.* (2011), is a commonly used technique for validating statistical regression models, particularly in cases where an additional, independent set of data is unavailable for validation.

In cross validation, the available data set – e.g. the 137 LLSOA data points described in section 2.3 – is split into a subset of 'training' data and a subset of 'test' data. The training data is used to generate the statistical regression model and estimated outputs for the quantities of interest – in this case, the Bayesian regression relationships, along with the posterior distributions for the energy intensities (e_j) and the area correction factors (β_j). This 'trained' statistical model is then used to make predictions for the test data points, and a comparison between the measured test data and the predictions at these data points is used to estimate the percentage error associated with the statistical model.

In Monte Carlo cross validation – also known as 'repeated random sub-sampling validation' – the cross validation process described above is repeated numerous times, with the overall data set split at random on each occasion into the training and test subsets using Monte Carlo sampling techniques. For each run of the cross validation, the percentage error is calculated between the measured test data and the predictions. The overall error is then estimated by averaging across the calculated errors for all of the Monte Carlo runs.

For this study, MATBUGS[15] – a Matlab-to-WinBUGS converter – was used to run the Monte Carlo cross validation of the Bayesian regression method, using 60 Monte Carlo runs. For each run, the 137 LLSOA data points – including the DECC measurements of energy consumption and the mapping data estimates of floor areas – were randomly split into a training data subset of 127 data points and a test data subset of 10 data points. In each run, the Bayesian regression method with error-in-variables, described in section 3.3, was used on the training data subset to generate posterior distributions for the energy intensity and area correction factor estimates.

Within each of the 60 Monte Carlo runs, a second (nested) Monte Carlo simulation was conducted, in which the LLSOA energy consumption (E_i) was predicted for each of the 10 test data points, using random samples from the posterior distributions for the energy intensities (e_j) and area correction factors (β_j), along with the floor area estimates (x_{ij}) and the relationships between these variables shown in Equations (1), (2) and (6). The percentage error was calculated for each run of the Monte Carlo cross validation, and the results were averaged across all the runs to estimate the overall error associated with the Bayesian regression method.

Using the Monte Carlo cross validation technique, it was seen that the overall percentage error for the Bayesian regression method was 4.1% – i.e. an underestimate in terms of the predicted LLSOA energy consumption in comparison to the DECC measurements. This 4.1% underestimate of district-level energy consumption from a statistical model is also smaller in magnitude than the estimates generated by the bottom-up engineering-based model demonstrated in Cheng and Steemers (2011), in which the predicted gas consumption, electricity consumption, and CO_2 emissions for the Manchester local authority[16] were seen to have a mean percentage error of 6.5%, 15.1%, and 9.5%, respectively.

Figure 12 shows the measured and predicted energy consumption (blue scatter plot) for each of the 60 Monte Carlo cross validation runs. These values are the total energy consumption for the 10 LLSOA test data points, which are chosen at random for each Monte Carlo run. In addition, lines are drawn to show an exact linear fit (solid line) between measurements and predictions – i.e. $y = x$ – and to show a $\pm 10\%$ deviation (dashed lines) – i.e. $y = 0.9x$ and $y = 1.1x$. It can be seen from Figure 12 that the majority of Monte Carlo runs result in an underestimate of the energy consumption, but that some runs result in a predicted energy consumption that is an overestimate or is very close to the measured energy consumption.

This underestimate between measured and predicted energy consumption is likely to be due to additional sources of uncertainty that have not been accounted for with the method outlined in section 3.3. For example, the assumption that the area correction factor (β_j) is consistent for each building class may be incorrect. In addition, the uncertainty associated with

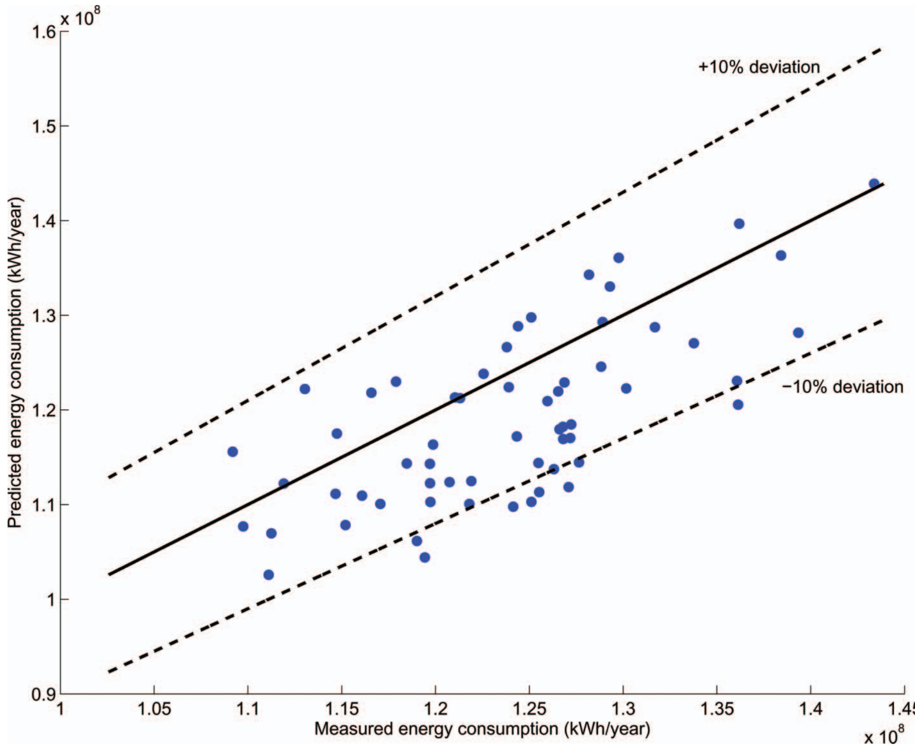

Figure 12. Measured and predicted energy consumption for the Monte Carlo cross validation.

the LLSOA energy consumption measurements (E_i) may not be fully captured by the relationships shown in Equation (3) and Figure 10.

Nevertheless, a percentage error of 4.1% is still relatively small compared to the uncertainty associated with the energy consumption data itself and is seen to be smaller than bottom-up engineering-based methods of estimating district-level energy consumption. The Bayesian regression analysis, therefore, can be considered to be accurate to within an acceptable level for the purpose of future applications. In particular, we can have confidence that the posterior distributions of energy intensity generated by the Bayesian regression analysis – which also quantify the uncertainty associated with these estimates – can be used to generate 'virtual observations' for the calibration of a bottom-up engineering-based housing stock model, as explained in section 4.2 below.

4.2. *Calibrating the housing stock model*

As discussed in sections 1 and 2.2, the ultimate aim of the analysis presented in this paper is to provide a hierarchical framework for calibrating a bottom-up engineering-based housing stock model – which models energy consumption at the micro-level (i.e. at the scale of an individual building) – when empirical energy data for calibration is only available at the macro-level (i.e. at a district or urban scale). The purpose of the Bayesian regression method, therefore, is to generate energy intensity estimates (in kWh/m^2/year) for each building class from the macro-level energy data, which can then be used to calibrate a bottom-up engineering-based housing stock model (such as SUSDEM) in the absence of building level empirical data.

Figure 13 describes the process by which the results of the Bayesian regression can be utilized in a Bayesian calibration to quantify the uncertainty in the input parameters, which can then be propagated through a bottom-up housing stock energy model to perform a retrofit analysis that is able to quantify the risks associated with certain technologies. This process consists of three steps:

(1) *Regression* – The Bayesian regression method, as described in this paper, is used to generate posterior distributions of the energy intensity for each building class. These posterior distributions are used to generate samples of energy intensity that act as 'virtual observations'.
(2) *Calibration* – The virtual observations (obtained by sampling from the posterior distributions generated by the Bayesian regression) are then used in a Bayesian calibration of a micro-level engineering-based model in order to quantify the uncertain model parameters, such as the internal heating set-point temperature and the infiltration rate, and thus improve the accuracy of the model. This calibration is conducted

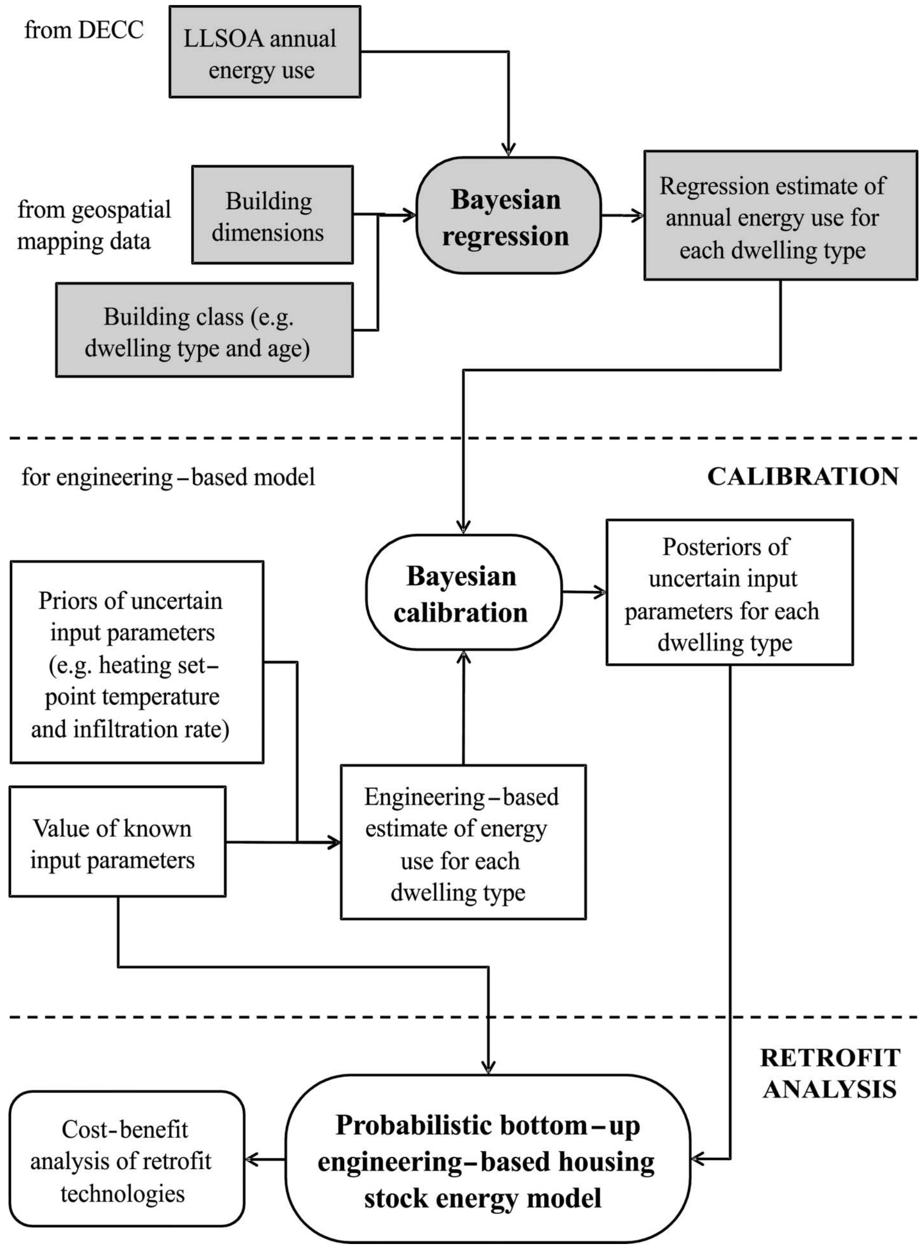

Figure 13. Process for calibrating a bottom-up engineering-based housing stock model.

separately for each building class. The Bayesian calibration works by assigning prior distributions to the uncertain input parameters, and then comparing the model energy predictions against observations in order to generate more accurate estimates for the uncertain parameters. The use of Bayesian calibration in the context of a bottom-up engineering-based housing stock model is explained and demonstrated for SUS-DEM in Booth *et al*. (2012).

(3) *Retrofit analysis* – The result of the Bayesian calibration is a set of posterior distributions for the uncertain input parameters of the model. This uncertainty can then be propagated through the bottom-up housing stock model using probabilistic sensitivity analysis (e.g. the Monte Carlo method) in order to quantify the uncertainty in the energy savings associated with different retrofit options. In combination with a probabilistic cost model, a full cost-benefit analysis can therefore be conducted that demonstrates the potential risks of different retrofit technologies to decision makers.

Figure 14 demonstrates the application of this hierarchical Bayesian calibration framework for the case of a single building class – in this case building class 7, the 1914–1945 semi-detached dwellings. The three sections of Figure 14 correspond to the three sections described above and shown in Figure 13: Bayesian regression to generate an estimate of the energy intensity for the building class, in the shape of a posterior distribution, which quantifies the uncertainty surrounding this estimate; Bayesian calibration, in which samples from the energy intensity posteriors are used to calibrate the uncertain input parameters in a bottom-up (micro-level) engineering-based domestic energy model; and retrofit analysis, whereby the uncertainty surround the input parameters is propagated through the bottom-up engineering-based model, in order to predict the energy demands associated with both baseline and retrofit scenarios, along with showing the uncertainty associated with these predictions.

In Figure 14, the top section – the Bayesian regression – is generated using the method described in section 3.3, and is identical to the graph for e_7 in Figure 11. The posterior distribution for the Bayesian regression estimate of the energy intensity, shown as a green dashed line in Figure 14, is sampled from in order to generate the 'virtual observations', which are then used in the second stage of the process – the Bayesian calibration.

In the Bayesian calibration – shown in the middle section of Figure 14 – prior distributions are given for the uncertain input parameters in the bottom-up engineering-based energy model. These prior distributions quantify the uncertainty associated with the average value of the given input parameter over the building class in question (e.g. the 1914–1945 semi-detached dwellings). In this case, six uncertain parameters were chosen for calibration, and the prior distributions for these parameters are shown as red (solid) lines in the middle section of Figure 14.

The Bayesian calibration process – described in Kennedy and O'Hagan (2001) and demonstrated for SUSDEM in Booth *et al.* (2012) – is applied, using the 'virtual observations' of energy intensity from the Bayesian regression and the prior distributions of the uncertain input parameters, in order to generate posterior distributions for these uncertain input parameters, as shown by the blue (dashed) lines in the middle section of Figure 14. Similarly to the prior distributions, these posteriors quantify the uncertainty associated with the average value of the given input parameter over the building class in question.[17]

The result is a calibrated (stochastic) bottom-up engineering-based energy model for the building class in question, which predicts the same energy intensity as given by the Bayesian regression analysis. This can be seen from the bottom section of Figure 14, which shows the predicted energy demands from the calibrated bottom-up engineering-based model for building class 7. It can be seen that the distribution of the baseline estimate of the average energy intensity across this building class (shown by the solid red line in the bottom section) is approximately equivalent to the posterior distribution of energy intensity given by the Bayesian regression (shown by the dashed green line in the top section). The mean and standard deviation of the posterior distribution from the Bayesian regression are 191 kWh/m^2/year and 14.2 kWh/m^2/year, respectively, compared to a mean and standard deviation for the estimate of energy intensity from the calibrated bottom-up engineering-based model for building class 7 of 189 kWh/m^2/year and 12.0 kWh/m^2/year. This difference in the mean energy intensity estimates between the Bayesian regression and the bottom-up engineering-based model amounts to a percentage error of only 1.0%, demonstrating that the hierarchical Bayesian framework provides a consistent method for calibrating a bottom-up engineering-based model.

It should be noted that, throughout this hierarchical calibration framework, any uncertainty in the earlier stages of the process will be propogated through to the later stages. For example, the spreads of the posterior distributions in Figure 11 indicate that there is still a great deal of uncertainty regarding the average energy intensity for each building class, and this uncertainty is propogated through to the Bayesian calibration stage. The calibration process works by comparing the outputs of the engineering-based model – the estimates of energy intensity for a given set of input parameters – against the 'observed' data, provided by samples from the posteriors of the Bayesian regression, in order to find the most likely values of the input parameters for the engineering-based model for these given 'observations'. If there is uncertainty surrounding these 'observations' – i.e. uncertainty in the distributions of energy intensity from the Bayesian regression – then this will be reflected by a greater uncertainty in the posterior distributions of the calibration parameters from the Bayesian calibration process, as shown in the middle section of Figure 14.

In turn, if there is a greater uncertainty surrounding the values of the calibration parameters, then this uncertainty will be propogated through into the retrofit analysis stage, and there will be a greater uncertainty surrounding the estimates of energy consumption from the engineering-based energy model, which will manifest itself in the form of a wider distribution for the predictions of energy consumption. The only way to reduce the propogation of such uncertainty through the framework is to reduce the uncertainty associated with the estimates of the energy intensities from the Bayesian regression stage. This can be done either by increasing the quantity and quality of

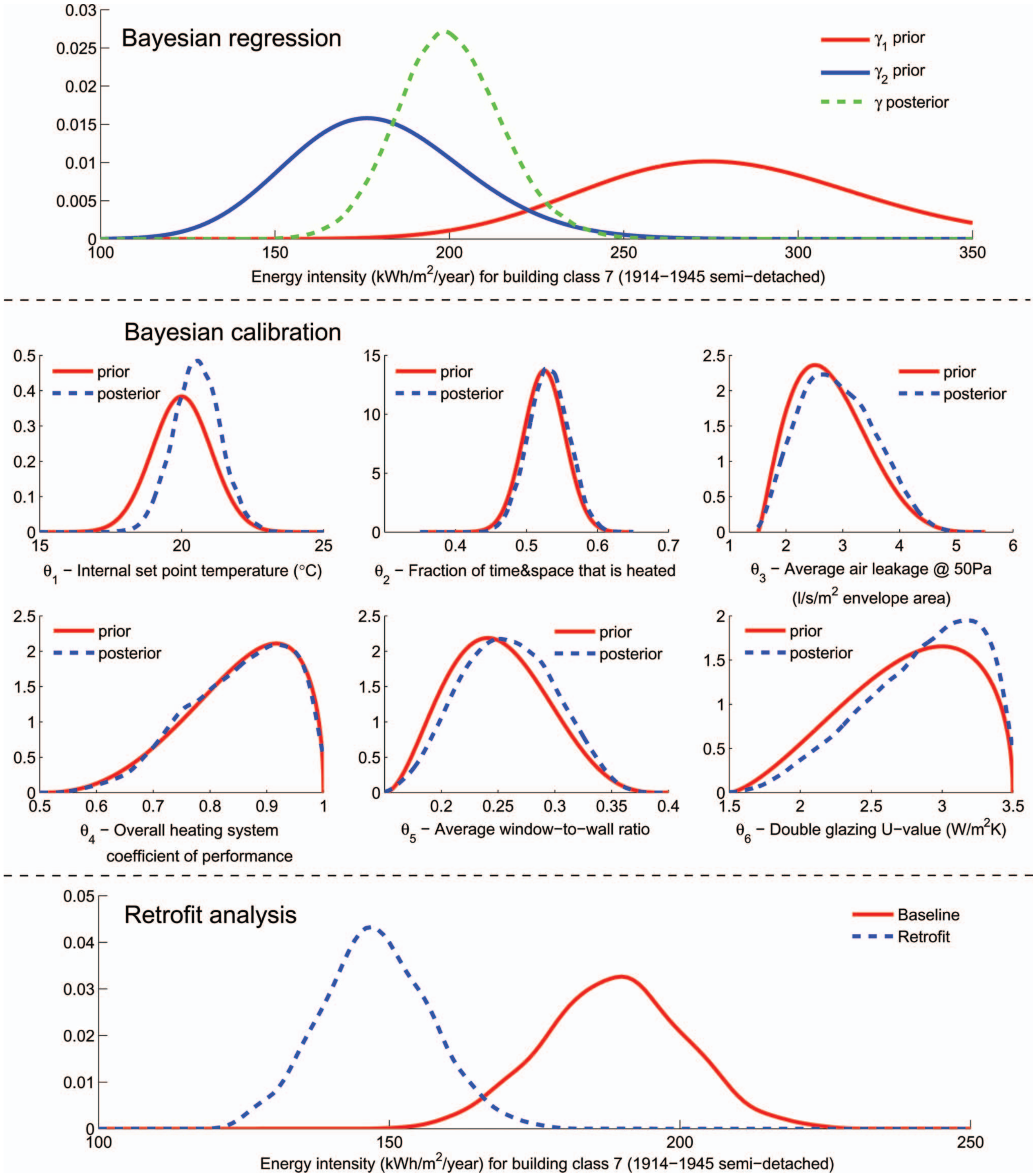

Figure 14. Demonstration of the hierarchical Bayesian calibration framework for building class 7 (1914–1945 semi-detached).

the macro-level data – i.e. the DECC LLSOA energy data – or by providing additional micro-level data to supplement the analysis – e.g. actual energy consumption data for individual dwellings for a representative sample of properties.

The calibrated bottom-up engineering-based model can also be used to predict the energy demands for a housing stock associated with retrofit scenarios, as shown by the dashed blue line in the bottom section of Figure 14. This distribution shows the average energy intensity estimated by the calibrated bottom-up engineering-based model for a retrofit scenario in which five retrofit measures[18] were applied simultaneously across 937 dwellings from building class 7 (1914–1945 semi-detached) in Salford. In addition, by propagating uncertainties through the model (including uncertainties surrounding the performance of the retrofit measures themselves), the uncertainty regarding this estimated retrofit energy demand is quantified. This allows for the risk associated with any planned retrofit intervention to be assessed by decision-makers. The authors plan to demonstrate the application of the calibrated bottom-up engineering-based housing stock energy model in relation to retrofit analysis and decision making in future work.

5. Conclusions

Those attempting to model domestic energy consumption or conduct retrofit analysis continually face the problem of a lack of energy consumption data at an appropriate temporal and spatial resolution, as has been recognized by Summerfield *et al*. (2011). This empirical data is required to improve the accuracy of energy models through calibration. In the case of retrofit analysis, using engineering-based bottom-up housing stock energy models, energy consumption data for a representative sample of the housing stock is difficult to obtain at the micro-level – i.e. from a broad sample of individual buildings. Empirical energy data is available, however, at a larger scale, such as at the district or urban level. In addition, there is a wide array of other energy data available in the form of previous studies and literature, benchmarks and guidelines, and expert opinion. Such prior information is rarely utilized, however, due to problems of scale or incompleteness.

This paper has presented a hierarchical framework that allows macro-level energy data to be used to generate energy intensity estimates, which in turn can be used as 'virtual observations' in order to calibrate the parameters in an engineering-based bottom-up housing stock energy model. Importantly, this framework also allows for prior information and empirical data to be integrated into a single analysis for calculating these energy intensity estimates.

A combination of Bayesian inference and regression analysis has been used to produce estimates of the average energy intensity for various dwelling types from empirical macro-level data. Unlike a standard linear regression, the use of Bayesian regression avoids nonsensical average energy intensity estimates (i.e. negative values), and allows subjective opinion and expert judgement to be incorporated into the analysis. In addition, the use of Bayesian methods enables the uncertainty over these average energy intensities to be quantified in the form posterior distributions, as well quantifying the uncertainty over the empirical data.

Limitations and uncertainties in existing sources of data have been discussed, whilst methods to accommodate for these within the Bayesian regression analysis have been proposed. The Bayesian regression method has been validated, and a hierarchical framework has been outlined and demonstrated for how to integrate Bayesian regression analysis with the Bayesian calibration of a bottom-up engineering-based housing stock model, such as SUSDEM. In this way, multiple sources of information at different scales can be integrated flexibly into a single retrofit analysis, depending on the data sources that are available to the user.

Finally, it should be noted that the Bayesian regression method demonstrated in this paper used publicly available macro-level energy consumption data, which can be obtained for any area of the UK and which is gathered every year. The analysis could therefore be greatly strengthened by utilizing the full breadth and depth of information that is available, instead of just using a small subset from a specific area and a single year. The real strength of employing Bayesian techniques lies in their ability to combine multiple sources of information in order to improve the accuracy of the predictions. The Bayesian regression method, therefore, is best thought of as a continual process in which newer, more reliable sources of information are integrated into the analysis as they become available in order to reduce the uncertainty of predictions over time.

Acknowledgement

Many thanks to City West Housing Trust in West Salford, UK, whose help and cooperation have been invaluable in this project. Thanks also to all the reviewers and their comments, which have been incredibly valuable for the authors.

Notes

1. In the UK, a large number of surveys of domestic buildings are conducted every year in order to produce Energy Performance Certificates (EPCs), which provide an official benchmark figure for the energy efficiency of the dwelling. Such surveys are used to collect data – including information on the age and structural type of a dwelling, its dimensions and number of rooms, the construction type of the walls, floors, roof, and glazing, and information on any heating systems present – which then act as inputs for the Reduced Data Standard Assessment Procedure (see BRE (2010) for details of RdSAP) in order to produce the EPCs. However, these surveys do not (as yet) include information about how the building is used – e.g. internal set point temperatures, number of occupants, number and type of appliances, behavioural patterns, etc.; as a result, the estimates produced by RdSAP and displayed on EPCs are only

notional, benchmark figures of energy consumption, relating to the energy efficiency of the building, and do not provide an accurate representation of the real energy consumption of the dwelling and its occupants. In addition, the EPC/RdSAP surveys do not gather information about the actual energy use of a household – e.g. from utility bills or meter readings.

2. Available from: www.decc.gov.uk/en/content/cms/statistics/regional/mlsoa_llsoa/mlsoa_llsoa.aspx.
3. Light Detection And Ranging.
4. Windows-run Bayesian inference Using Gibbs Sampling (Available from: www.mrc-bsu.cam.ac.uk/bugs/).
5. Surveys of nearly 6300 sample properties from the Salford housing stock – conducted in order to generate EPCs – indicate that approximately 5.5% of dwellings use electricity as their primary source of fuel for space heating and hot water, with a further 6.6% of dwellings using electricity as a secondary source of space heating (e.g. a direct electric heater) alongside a gas boiler. It is therefore not possible to separate the fuel consumption data – for gas and electricity – into different end-uses; i.e. gas consumption figures are not equivalent to the energy demand for space heating and hot water, whilst electricity consumption figures are not equivalent to the energy demand for lighting and appliances. It is for this reason that the gas and electricity consumption data has been combined to give the total energy consumption for each LLSOA.
6. Mapping data was obtained from UKMap by The GeoInformation Group (Available from: www.geoinformationgroup.co.uk/products/mapping)
7. The additional uncertainty that is introduced by removing 0.6% of the dwelling floor area from the calculation (whilst the energy consumption data remains unchanged) is negligible compared to the various other sources of uncertainty involved in the analysis; for example, the uncertainty concerned with calculating the dwelling floor areas from the mapping data or with the energy consumption data itself (as discussed in later sections). In addition, even if one is able to eliminate all other sources of uncertainty, and if one assumes a linear relationship between energy consumption and floor area, a change to the floor area of less than 1% will cause a change in the calculated energy intensity of less than 1%, which is also considered negligible for the purpose of this study. It is up to the analyst or decision-maker to decide what is an acceptable tolerance of accuracy, and in other cases this may be deemed to be an intolerable level of accuracy.
8. Measurements of internal floor areas – available for a large sample of properties in the case study area from surveys conducted for EPCs – show that floor areas are, on average, consistent in multi-storey dwellings – i.e. that the ground floor (footprint) floor area is the same as that of the floors above.
9. Reports CE189 and CE190 from the Energy Saving Trust (Available from: www.energysavingtrust.org.uk).
10. The CEN–ISO standards are a calculation methodology outlined by the European Committee for Standardization (CEN) and the International Organization for Standardization (ISO).
11. These assumptions are outlined in detail in the Standard Assessment Procedure, developed by the BRE (2010).
12. $f(x) = \frac{1}{B(\alpha_1,\alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}$, where B is the Beta function.
13. $f(x) = \frac{x^{a-1}}{b^a \Gamma(a)} exp(\frac{-x}{b})$, where Γ is the Gamma function.
14. In addition, in order to allow for the error-in-variables model, the energy consumption data (E_i) and the floor areas (x_{ij}) were no longer normalized by the total floor areas for each LLSOA ($\sum_{j=1}^{P} x_{ij}$) – as was previously explained in section 2.2 – since these normalization values are now themselves treated as uncertain quantities.
15. Available from: http://code.google.com/p/matbugs/
16. A local authority (LA) is an administrative area, composed of many LLSOAs. For example, Salford is a LA in the Greater Manchester area in the north-west of England composed of 144 LLSOAs. Manchester is another LA in the Greater Manchester area.
17. It should be noted, however, that although the average value of the input parameters is altered uniformly by the Bayesian calibration for all dwellings within the building class, variation (i.e. heterogeneity) between individual dwellings in terms of all input parameters is still accounted for by the stochastic nature of SUSDEM, which uses second-order Monte Carlo simulation, as described in Halpern *et al*. (2000) and Booth *et al*. (2012).
18. The five measured were cavity wall insulation, double glazing, loft insulation, condensing boilers and draught-proofing.

References

Arlot, S., 2010. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40–79.

Booth, A., Choudhary, R., and Spiegelhalter, D., 2012. Handling uncertainty in housing stock models. *Building and Environment*, 48, 35–47.

BRE, 2009 [2010]. *The governments standard assessment procedure for energy rating of dwellings*. Building Research Establishment. Technical report SAP 2009 version 9.90.

Bruhns, H. and Wyatt, P., 2011. A data framework for measuring the energy consumption of the non-domestic building stock. *Building Research Information*, 39 (3), 211–226.

Caves, D.W., *et al*., 1987. A Bayesian approach to combining conditional demand and engineering models of electricity usage. *The Review of Economics and Statistics*, 69 (3), 438–448.

Cheng, V. and Steemers, K., 2011. Modelling domestic energy consumption at district scale: a tool to support national and local energy policies. *Environmental Modelling & Software*, 26 (10), 1186–1198.

DECC, 2010. *Guidance note for regional energy data*. [online]. Department of Energy and Climate Change. Technical report 10D/1003. Available from: http://www.decc.gov.uk/assets/decc/Statistics/regional/1087-guidance-note-regional-energy-data.pdf [Accessed 11 September 2012].

DECC, 2011. *Guidance note for the DECC MLSOA/IGZ and LLSOA electricity and gas consumption data*. [online]. Department of Energy and Climate Change. Technical report 11D/671. Available from: http://www.decc.gov.uk/assets/decc/statistics/regional/mlsoa2008/1_20100325121429_e_@@_mlsoallsoaguidance.pdf [Accessed 11 September 2012].

Dellaportas, P. and Stephens, D.A., 1995. Bayesian analysis of errors-in-variables regression models. *Biometrics*, 51 (3), 1085–1095.

Firth, S.K., Lomas, K.J., and Wright, A.J., 2010. Targeting household energy-efficiency measures using sensitivity analysis. *Building Research & Information*, 38 (1), 25.

Halpern, E.F., *et al.*, 2000. Representing both first- and second-order uncertainties by Monte Carlo simulation for groups of patients. *Medical Decision Making*, 20 (3), 314–322.

Heo, Y., Choudhary, R., and Augenbroe, G., 2012. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47, 550–560.

Jackson, C., Best, N., and Richardson, S., 2006. Improving ecological inference using individual level data. *Statistics in Medicine*, 25 (12), 2136–2159.

Jackson, C.H., Best, N.G., and Richardson, S., 2009. Bayesian graphical models for regression on multiple data sets with different variables. *Biostatistics*, 10 (2), 335–351.

Kavgic, M., *et al.*, 2010. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, 45 (7), 1683–1697.

Kennedy, M.C. and O'Hagan, A., 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63 (3), 425–464.

Lomas, K.J. and Eppel, H., 1992. Sensitivity analysis techniques for building thermal simulation programs. *Energy and Buildings*, 19 (1), 21–44.

Lunn, D.J., *et al.*, 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10 (4), 325–337.

Macdonald, I., 2002. *Quantifying the effects of uncertainty in building simulation*. Thesis (PhD). University of Strathclyde.

Milne, G. and Boardman, B., 2000. Making cold homes warmer: the effect of energy efficiency improvements in low-income homes: a report to the Energy Action Grants Agency Charitable Trust. *Energy Policy*, 28 (6–7), 411–424.

O'Hagan, A., 2006. *Uncertain judgements: eliciting experts' probabilities*. John Wiley and Sons.

O'Hagan, A. and Oakley, J.E., 2004. Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering & System Safety*, 85 (1–3), 239–248.

Oreszczyn, T., *et al.*, 2006. Determinants of winter indoor temperatures in low income households in England. *Energy and Buildings*, 38 (3), 245–252.

Spiegelhalter, D.J. and Best, N.G., 2003. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine*, 22 (23), 3687–3709.

Spiegelhalter, D.J. and Riesch, H., 2011. Don't know, can't know: embracing deeper uncertainties when analysing risks. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 369 (1956), 4730–4750. PMID: 2204–2895.

Summerfield, A.J., *et al.*, 2011. How useful are building energy models for policy? A UK perspective. *In*: *12th conference of international building performance simulation association*, 14–16 November, IBPSA, Sydney, Australia.

Swan, L.G. and Ugursal, V.I., 2009. Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13 (8), 1819–1835.

Witten, I.H., Frank, E., and Hall, M.A., 2011. *Data mining: practical machine learning tools and techniques*. Elsevier.